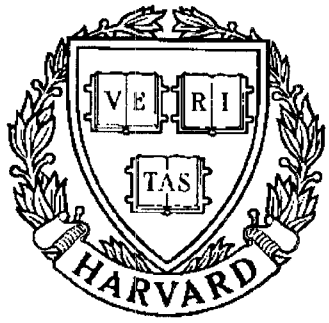
TECHNICAL RESEARCH REPORT

SYSTEMS RESEARCH CENTER



Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), Industry and the University

Estimation of Multiple Sinusoids by Parametric Filtering

by T-H. Li and B. Kedem

TR 92-51

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JAN 1992		**00-00-1992 to 00-00-1992**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Estimation of Multiple Sinusoids by Parametric Filtering	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Maryland,Systems Research Center,College Park,MD,20742**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**50**	

ESTIMATION OF MULTIPLE SINUSOIDS BY PARAMETRIC FILTERING[1]

Ta-Hsin Li and Benjamin Kedem [2]

January 1992

[1]This work was supported by grants AFOSR-89-0049, ONR N00014-89-J-1051, and NSF CDR-88-03012.

[2]The authors are with the Department of Mathematics and the Systems Research Center, University of Maryland, College Park, Maryland 20742.

Abstract

The problem of estimating the frequencies of multiple sinusoids from noisy observations is addressed in this paper. A parametric filtering approach, called the PF method, is proposed that leads to a consistent estimator of the AR representation of the sinusoidal signal, given the number of sinusoids. It is accomplished by using an iterative procedure to find a fixed-point of the parametrized least squares estimator (from the filtered data) that comprises a contraction mapping in the vicinity of the true AR parameter. Employing appropriate filters, this method is able to achieve the accuracy of the nonlinear least squares estimator, with much less computational complexity and initialization requirement. It can also be implemented adaptively (recursively) in order to track time-varying frequencies. In this way, the PF method provides a flexible and efficient procedure of frequency estimation. An example of the AR filter is investigated in detail to illustrate the performance of the PF method.

Abbreviated Title: "Frequency Estimation by Parametric Filtering"

Key words and phrases: Autoregressive, contraction mapping, fixed-point, frequency estimation, iterative filtering, least squares, parametric filtering, sinusoid, spectral analysis.

1 Introduction

Let $\{x_t\}$ be a random process (signal) consisting of q superimposed real sinusoids with unknown frequencies, as given by

$$x_t := \sum_{k=1}^{q} \beta_k \cos(\omega_k t + \phi_k). \tag{1.1}$$

In this expression, $q > 0$ is a known integer, the β_k and ω_k are unknown constants, satisfying $\beta_k > 0$ and $0 < \omega_1 < \cdots < \omega_q < \pi$, and the ϕ_k are independent and uniformly distributed random variables on $[0, 2\pi)$. Suppose that the signal is contaminated by additive noise so that the observed mixed-spectrum process, denoted by $\{y_t\}$, can be written as

$$y_t := x_t + \epsilon_t \qquad (t = 0, \pm 1, \pm 2, \dots). \tag{1.2}$$

Assume that the noise $\{\epsilon_t\}$ is independent of $\{\phi_k\}$ and can be modeled as a linear process of the form

$$\epsilon_t := \sum_{j=-\infty}^{\infty} \psi_j \, \xi_{t-j}, \qquad \{\xi_t\} \sim \mathrm{IID}(0, \sigma_\xi^2), \;\; \sum |\psi_j| < \infty. \tag{1.3}$$

Under these assumptions, $\{y_t\}$ becomes stationary, with mean zero and autocovariance function

$$r_\tau^y := E(y_{t+\tau} y_t) = r_\tau^x + r_\tau^\epsilon,$$

where $r_\tau^x := E(x_{t+\tau} x_t)$ and $r_\tau^\epsilon := E(\epsilon_{t+\tau} \epsilon_t)$ are autocovariance functions of the signal and the noise, respectively, and can be written as

$$r_\tau^x = \sum_{k=1}^{q} \tfrac{1}{2} \beta_k^2 \cos(\omega_k \tau) \quad \text{and} \quad r_\tau^\epsilon = \sigma_\xi^2 \sum_j \psi_{j+\tau} \psi_j. \tag{1.4}$$

The objective of frequency estimation is to find estimators of the sinusoidal frequencies ω_k on the basis of a finite data set $\{y_1, \dots, y_n\}$ obtained from the random process (1.2).

Among a variety of different procedures of frequency estimation in the literature, there is one, known as the AR approach (Kay and Marple, 1981), that resorts to the AR

representation of the sinusoidal signal under investigation. In fact, $\{x_t\}$ is a solution of the following homogeneous autoregressive (AR) equation of order $2q$:

$$\sum_{j=0}^{2q} a_j \, x_{t-j} = 0. \tag{1.5}$$

The AR coefficients a_j in this equation are defined as the coefficients of the polynomial

$$A(z) := \sum_{j=0}^{2q} a_j \, z^{2q-j} := \prod_{k=1}^{q} (z - e^{i\omega_k})(z - e^{-i\omega_k}). \tag{1.6}$$

It is clear that the a_j in (1.5) completely determined by ω_k, and vice versa. Moreover, the a_j are real and symmetric in the sense that

$$a_0 = 1 \quad \text{and} \quad a_{2q-j} = a_j \quad (j = 0, \ldots, q-1). \tag{1.7}$$

Therefore, the original problem of frequency estimation can be equivalently stated as that of estimating the AR parameter vector $\mathbf{a} := [a_1, \cdots, a_q]^T$, where the superscript T denotes matrix transposition. This *reparametrization* enables us to employ many well-studied *linear* methods, which usually end up with solving systems of linear equations. Once an estimate of $\mathbf{a}$ becomes available, the frequency estimates can be obtained from the zeros of $A(z)$ in (1.6), with a_j replaced by their estimates.

A widely-used procedure for estimating the AR parameter $\mathbf{a}$ is the so-called least squares (LS) estimator, also known as Prony's spectral line estimator (Kay and Marple, 1981), which can be summarized as follows. Due to the AR model (1.5), the observed process $\{y_t\}$ satisfies the equation

$$\sum_{j=0}^{2q} a_j \, y_{t-j} = e_t \quad \text{with} \quad e_t := \sum_{j=0}^{2q} a_j \, \epsilon_{t-j}. \tag{1.8}$$

For a given time series $\{y_1, \ldots, y_n\}$ of length $n > 2q$, (1.8), together with (1.7), gives

$$\mathbf{y} = -\mathbf{Y}\mathbf{Q}\mathbf{a} + \mathbf{e}. \tag{1.9}$$

In this equation,

$$\mathbf{y} := \begin{bmatrix} y_{2q+1} + y_1 \\ \vdots \\ y_n + y_{n-2q} \end{bmatrix} \quad \mathbf{Y} := \begin{bmatrix} y_{2q} & \cdots & y_2 \\ \vdots & & \vdots \\ y_{n-1} & \cdots & y_{n-2q+1} \end{bmatrix} \quad \mathbf{e} := \begin{bmatrix} e_{2q+1} \\ \vdots \\ e_n \end{bmatrix} \tag{1.10}$$

and $\mathbf{Q}$ is a $(2q-1)$-by-q matrix of the form

$$\mathbf{Q} := \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}^T & 1 \\ \tilde{\mathbf{I}} & \mathbf{0} \end{bmatrix}$$

where $\mathbf{I}$ stands for the $(q-1)$-by-$(q-1)$ identity matrix, $\tilde{\mathbf{I}}$ the $(q-1)$-by-$(q-1)$ reverse permutation matrix, with 1's on the antidiagonal and 0's elsewhere, and $\mathbf{0}$ the zero vector of dimension $q-1$. The LS estimator of $\mathbf{a}$, denoted by $\mathbf{a}_{\mathrm{LS}}$, is defined as the minimizer of the sum of squared errors $\|\mathbf{y} + \mathbf{YQa}\|^2$. Simple algebra gives

$$\mathbf{a}_{\mathrm{LS}} = -\mathbf{Y}^{\dagger}\mathbf{y} \tag{1.11}$$

where $\mathbf{Y}^{\dagger}$ is the pseudo-inverse of $\mathbf{YQ}$, as specified by

$$\mathbf{Y}^{\dagger} := (\mathbf{Q}^T\mathbf{Y}^T\mathbf{YQ})^{-1}\mathbf{Q}^T\mathbf{Y}^T. \tag{1.12}$$

The AR approach is computationally attractive as compared to the nonlinear least squares (NLS) approach, since the former solves a system of linear equations, whereas the latter fits the data with a sum of sinusoids whose amplitudes and frequences are free variables, yielding a nonlinear optimization problem. On the other hand, the AR approach has also some serious shortcomings, among which is the *inconsistency* of $\mathbf{a}_{\mathrm{LS}}$ as an estimator of the AR parameter $\mathbf{a}$. In fact, as n tends to infinity, the strong ergodicity of $\{y_t\}$ gives

$$n^{-1}\mathbf{Y}^T\mathbf{Y} \overset{a.s.}{\to} \bar{\mathbf{R}}_y \quad \text{and} \quad n^{-1}\mathbf{Y}^T\mathbf{y} \overset{a.s.}{\to} \bar{\mathbf{r}}_y + \bar{\mathbf{r}}_y^B$$

(see Li and Kedem, 1991), where $\bar{\mathbf{R}}_y$ and $\bar{\mathbf{r}}_y$ are defined by

$$\bar{\mathbf{R}}_y := \begin{bmatrix} r_0^y & r_{-1}^y & \cdots & r_{-2q+2}^y \\ r_1^y & r_0^y & \cdots & r_{-2q+3}^y \\ \vdots & \vdots & \ddots & \vdots \\ r_{2q-2}^y & r_{2q-3}^y & \cdots & r_0^y \end{bmatrix} \quad \text{and} \quad \bar{\mathbf{r}}_y := \begin{bmatrix} r_1^y \\ r_2^y \\ \vdots \\ r_{2q-1}^y \end{bmatrix}, \tag{1.13}$$

respectively, with the superscript B standing for the backward rearrangement of a vector. From (1.9), we can also write

$$\begin{aligned}\bar{\mathbf{r}}_y + \bar{\mathbf{r}}_y^B &= (n-2q)^{-1}E(\mathbf{Y}^T\mathbf{y}) \\ &= (n-2q)^{-1}\{-E(\mathbf{Y}^T\mathbf{Y})\mathbf{Q}\mathbf{a} + E(\mathbf{Y}^T\mathbf{e})\} \\ &= -\bar{\mathbf{R}}_y\mathbf{Q}\mathbf{a} + \bar{\mathbf{r}}_\epsilon + \bar{\mathbf{r}}_\epsilon^B + \bar{\mathbf{R}}_\epsilon\mathbf{Q}\mathbf{a},\end{aligned}$$

where $\bar{\mathbf{R}}_\epsilon$ and $\bar{\mathbf{r}}_\epsilon$ are similarly defined from $\{\epsilon_t\}$ as $\bar{\mathbf{R}}_y$ and $\bar{\mathbf{r}}_y$. The last equality employs the identity $(n-2q)^{-1}E(\mathbf{Y}^T\mathbf{e}) = \bar{\mathbf{r}}_\epsilon + \bar{\mathbf{r}}_\epsilon^B + \bar{\mathbf{R}}_\epsilon\mathbf{Q}\mathbf{a}$, which can be easily verified upon noting that the vector $\mathbf{e}$ has a similar representation as (1.9), with $\mathbf{Y}$ and $\mathbf{y}$ replaced by the corresponding matrices from $\{\epsilon_t\}$. Using these results in (1.11) as $n \to \infty$ yields

$$\mathbf{a}_{\mathrm{LS}} \overset{a.s.}{\to} -\mathbf{R}_y^{-1}\mathbf{r}_y = \mathbf{a} - \mathbf{R}_y^{-1}(\mathbf{r}_\epsilon + \mathbf{R}_\epsilon\mathbf{a}), \tag{1.14}$$

where

$$\mathbf{R}_y := \mathbf{Q}^T\bar{\mathbf{R}}_y\mathbf{Q}, \quad \mathbf{r}_y := \mathbf{Q}^T(\bar{\mathbf{r}}_y + \bar{\mathbf{r}}_y^B) = 2\,\mathbf{Q}^T\bar{\mathbf{r}}_y,$$

and $\mathbf{R}_\epsilon$ and $\mathbf{r}_\epsilon$ are defined analogously from $\{\epsilon_t\}$. It is readily seen from (1.14) that the limit of $\mathbf{a}_{\mathrm{LS}}$ differs from the parameter $\mathbf{a}$ being estimated, since in general $\mathbf{r}_\epsilon + \mathbf{R}_\epsilon\mathbf{a} \neq \mathbf{0}$. In other words, the above AR approach produces an *inconsistent* estimator of the AR parameter $\mathbf{a}$. It is not surprising, since $\{e_t\}$ in (1.8) is not white so that $\{y_t\}$ obeys an ARMA rather than an AR model. As will be seen later, this problem of inconsistency can be resolved by a parametric filtering approach, which we call the parametric filtering (PF) method.

2 Parametric Filtering Method

To alleviate the inconsistency problem of the AR approach, we shall consider estimating the AR parameter $\mathbf{a}$, not directly from the original data, but from the *filtered* data, upon using an appropriate parametric filter. It will be shown that the LS estimator from the filtered data can be made consistent, by "tuning" the filter parameter so as to eliminate, in the filtered data, the counterpart of the vector $\mathbf{r}_\epsilon + \mathbf{R}_\epsilon\mathbf{a}$ in the limit.

2.1 Definition and Algorithm

Keeping this idea in mind, let us first introduce a linear time-invariant causal filter, whose impulse response sequence is denoted by $\{h_j(\boldsymbol{\alpha}), j = 0, 1, \dots\}$, where $\boldsymbol{\alpha} := [\alpha_1, \cdots, \alpha_q]^T$ is a parameter vector that takes on values in a closed and bounded subset $\mathcal{A}$ of a q-dimensional Euclidean space. Assume that the filter is BIBO-stable, that is, $\sum |h_j(\boldsymbol{\alpha})| < \infty$, for all $\boldsymbol{\alpha} \in \mathcal{A}$. Let $\{y_t(\boldsymbol{\alpha})\}$ be the filtered process given by

$$y_t(\boldsymbol{\alpha}) := \sum_{j=0}^{\infty} h_j(\boldsymbol{\alpha})\, y_{t-j},$$

and define the filtered signal $\{x_t(\boldsymbol{\alpha})\}$ and the filtered noise $\{\epsilon_t(\boldsymbol{\alpha})\}$ analogously.

According to the theory of linear filtering (see, e.g., Brockwell and Davis, 1987), the filtered signal $\{x_t(\boldsymbol{\alpha})\}$ remains a sum of q sinusoids which possess the *same* frequencies as the unfiltered signal $\{x_t\}$ in (1.1), with possibly different amplitudes and phases. As a matter of fact, $x_t(\boldsymbol{\alpha})$ can be expressed as

$$x_t(\boldsymbol{\alpha}) = \sum_{k=1}^{q} \beta_k(\boldsymbol{\alpha})\, \cos\{\omega_k t + \phi_k(\boldsymbol{\alpha})\},$$

where $\beta_k(\boldsymbol{\alpha})$ and $\phi_k(\boldsymbol{\alpha})$ are amplitudes and phases of the filtered sinusoids, as specified by

$$\beta_k(\boldsymbol{\alpha}) := \beta_k\, |H(\omega_k; \boldsymbol{\alpha})| \quad \text{and} \quad \phi_k(\boldsymbol{\alpha}) := \phi_k + \angle H(\omega_k; \boldsymbol{\alpha}) \mod 2\pi,$$

respectively, and $H(\omega; \boldsymbol{\alpha})$ is the transfer function of the filter, as defined by

$$H(\omega; \boldsymbol{\alpha}) := \sum_{j=0}^{\infty} h_j(\boldsymbol{\alpha})\, e^{-ij\omega}.$$

It is quite clear that the original problem of frequency estimation is not altered by the linear filtering and can be equivalently stated as that of estimating the *same* AR parameter $\mathbf{a}$ in terms of the *filtered* processes.

Now let $\mathbf{a}_{\mathrm{LS}}(\boldsymbol{\alpha})$ be the LS estimator of $\mathbf{a}$ on the basis of the filtered time series $\{y_1(\boldsymbol{\alpha}), \dots, y_n(\boldsymbol{\alpha})\}$. Since the strong ergodicity still holds for $\{y_t(\boldsymbol{\alpha})\}$, it can be shown as before that $\mathbf{a}_{\mathrm{LS}}(\boldsymbol{\alpha})$ converges almost surely to the vector

$$\mathbf{a}(\boldsymbol{\alpha}) := -\mathbf{R}_y^{-1}(\boldsymbol{\alpha})\, \mathbf{r}_y(\boldsymbol{\alpha}) \tag{2.1}$$

which, similar to (1.14), can be written as

$$\mathbf{a}(\boldsymbol{\alpha}) = \mathbf{a} - \mathbf{R}_y^{-1}(\boldsymbol{\alpha})\,\{\mathbf{r}_\epsilon(\boldsymbol{\alpha}) + \mathbf{R}_\epsilon(\boldsymbol{\alpha})\,\mathbf{a}\}. \tag{2.2}$$

In these expressions, the autocovariances are defined from the corresponding filtered processes. Multiplying each side of (2.2) by $\mathbf{R}_y(\boldsymbol{\alpha})$, and since $\mathbf{R}_y(\boldsymbol{\alpha}) = \mathbf{R}_x(\boldsymbol{\alpha}) + \mathbf{R}_\epsilon(\boldsymbol{\alpha})$, we obtain

$$\mathbf{R}_x(\boldsymbol{\alpha})\,\mathbf{a}(\boldsymbol{\alpha}) + \mathbf{R}_\epsilon(\boldsymbol{\alpha})\,\mathbf{a}(\boldsymbol{\alpha}) = \mathbf{R}_x(\boldsymbol{\alpha})\,\mathbf{a} - \mathbf{r}_\epsilon(\boldsymbol{\alpha}).$$

Assuming for the time being that $\mathbf{R}_x(\boldsymbol{\alpha})$ is nonsingular, we obtain

$$\mathbf{a} - \mathbf{a}(\boldsymbol{\alpha}) = \mathbf{R}_x^{-1}(\boldsymbol{\alpha})\,\{\mathbf{r}_\epsilon(\boldsymbol{\alpha}) + \mathbf{R}_\epsilon(\boldsymbol{\alpha})\,\mathbf{a}(\boldsymbol{\alpha})\}. \tag{2.3}$$

Evidently, if there exists a filter parameter $\boldsymbol{\alpha}^*$ in the interior of $\mathcal{A}$ such that $\mathbf{R}_\epsilon(\boldsymbol{\alpha}^*)$ is nonsingular and

$$\mathbf{a}(\boldsymbol{\alpha}^*) = -\mathbf{R}_\epsilon^{-1}(\boldsymbol{\alpha}^*)\,\mathbf{r}_\epsilon(\boldsymbol{\alpha}^*), \tag{2.4}$$

then from (2.3) we would obtain $\mathbf{a}(\boldsymbol{\alpha}^*) = \mathbf{a}$. This implies that by fixing the filter parameter $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$, the corresponding LS estimator $\mathbf{a}_{\mathrm{LS}}(\boldsymbol{\alpha}^*)$ becomes consistent for estimating the AR parameter $\mathbf{a}$. In other words, the requirement (2.4) implies $\mathbf{r}_\epsilon(\boldsymbol{\alpha}^*) + \mathbf{R}_\epsilon(\boldsymbol{\alpha}^*)\,\mathbf{a} = \mathbf{r}_\epsilon(\boldsymbol{\alpha}^*) + \mathbf{R}_\epsilon(\boldsymbol{\alpha}^*)\,\mathbf{a}(\boldsymbol{\alpha}^*) = \mathbf{0}$.

Suppose that the autocorrelation function of the noise $\{\epsilon_t\}$ is known, and that the filter is parametrized so that

(A1) $$\boldsymbol{\alpha} = -\mathbf{R}_c^{-1}(\boldsymbol{\alpha})\,\mathbf{r}_\epsilon(\boldsymbol{\alpha}) \qquad \text{for all } \boldsymbol{\alpha} \in \mathcal{A}.$$

Under this assumption, (2.4) reduces to $\mathbf{a}(\boldsymbol{\alpha}^*) = \boldsymbol{\alpha}^*$, which indicates that the desired filter parameter $\boldsymbol{\alpha}^*$ is a *fixed-point* of the mapping $\mathbf{a}(\boldsymbol{\alpha})$. Moreover, since $\mathbf{a}(\boldsymbol{\alpha}^*) = \mathbf{a}$, it also follows that $\boldsymbol{\alpha}^* = \mathbf{a}$. Therefore, with (A1) being satisfied, it is no longer necessary to distinguish the AR parameter $\mathbf{a}$ and the desired filter parameter $\boldsymbol{\alpha}^*$, and the problem of estimating $\mathbf{a}$ becomes identical to that of estimating $\boldsymbol{\alpha}^*$.

It is interesting to note the similarity between (A1) and the Yule-Walker equations. In fact, for an AR($2q$) process with AR parameters satisfying (1.7), the vector of the first

q free parameters will be the solution of (A1), with $\mathbf{R}_\epsilon(\boldsymbol{\alpha})$ and $\mathbf{r}_\epsilon(\boldsymbol{\alpha})$ being replaced by their counterparts defined from the autocovariances of that AR process.

Given a *finite* sample of $\{y_t\}$, let $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ be a consistent estimator of $\mathbf{a}(\boldsymbol{\alpha})$. Since $\boldsymbol{\alpha}^*$ is a fixed-point of $\mathbf{a}(\boldsymbol{\alpha})$, it is reasonable to estimate $\boldsymbol{\alpha}^*\,(=\mathbf{a})$ by a fixed-point, denoted by $\hat{\boldsymbol{\alpha}}$, of the *random* mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$, so that

$$\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}}) = \hat{\boldsymbol{\alpha}}. \tag{2.5}$$

We refer to this procedure of frequency estimation as *parametric filtering method*, or simply PF method. The fixed-point $\hat{\boldsymbol{\alpha}}$ is called PF estimator of the AR parameter $\mathbf{a}\,(=\boldsymbol{\alpha}^*)$ and the corresponding angular frequencies, denoted by $\hat{\omega}_k$, of the zeros of $A(z)$ in (1.6), with $\hat{\boldsymbol{\alpha}}$ in place of $\mathbf{a}$, are called PF estimators of ω_k.

In order to find the PF estimator $\hat{\boldsymbol{\alpha}}$, a simple iterative algorithm, known as the *fixed-point iteration* (FPI), can be employed. This algorithm generates a sequence $\{\hat{\boldsymbol{\alpha}}_m\}$ of estimators according to the iterative procedure

$$\hat{\boldsymbol{\alpha}}_m = \hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}}_{m-1}) \qquad (m = 1, 2, \ldots). \tag{2.6}$$

It will be shown later that under certain conditions $\hat{\boldsymbol{\alpha}}_m$ converges to $\hat{\boldsymbol{\alpha}}$ as $m \to \infty$ almost surely, provided that the sample size n is sufficiently large.

Note that (2.6) can be regarded as an iterative filtering procedure, which can be implemented in two steps within each iteration:

STEP 1. Filter the data with the filter parameter $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}_{m-1}$;

STEP 2. Compute $\hat{\boldsymbol{\alpha}}_m = \hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}}_{m-1})$ from the filtered data.

Other iterative filtering procedures for frequency estimation also exist in the literature (see, e.g., in Kay, 1984; Dragošević and Stanković, 1989).

2.2 Least Squares Estimator

So far we have not specified the estimator $\hat{\mathbf{a}}(\boldsymbol{\alpha})$. In fact, any estimator would qualify as long as it converges (stochastically) to $\mathbf{a}(\boldsymbol{\alpha})$ when the sample size n tends to infinity. In

the following, we specialize the choice of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ by considering the LS estimator from the filtered data.

Notice that given the finite data $\{y_1, \ldots, y_n\}$, the filtered process $\{y_t(\boldsymbol{\alpha})\}$ can only be approximated by

$$\hat{y}_t(\boldsymbol{\alpha}) := \sum_{j=0}^{t-1} h_j(\boldsymbol{\alpha})\, y_{t-j} \qquad (t = 1, \ldots, n). \tag{2.7}$$

Therefore, the estimator $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ should be obtained from $\{\hat{y}_1(\boldsymbol{\alpha}), \ldots, \hat{y}_n(\boldsymbol{\alpha})\}$, rather than the unavailable data $\{y_1(\boldsymbol{\alpha}), \ldots, y_n(\boldsymbol{\alpha})\}$. Enlightened by the estimator $\mathbf{a}_{\mathrm{LS}}(\boldsymbol{\alpha})$, which depends on $\{y_t(\boldsymbol{\alpha})\}$, we may take

$$\hat{\mathbf{a}}(\boldsymbol{\alpha}) := -\hat{\mathbf{Y}}^{\dagger}(\boldsymbol{\alpha})\, \hat{\mathbf{y}}(\boldsymbol{\alpha}), \tag{2.8}$$

where $\hat{\mathbf{y}}(\boldsymbol{\alpha})$ and $\hat{\mathbf{Y}}^{\dagger}(\boldsymbol{\alpha})$, defined from $\{\hat{y}_t(\boldsymbol{\alpha})\}$, are the counterparts of $\mathbf{y}$ and $\mathbf{Y}^{\dagger}$ given by (1.10) and (1.12). It is readily shown that $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ in (2.8) is the LS estimator that minimizes the criterion $\|\hat{\mathbf{y}}(\boldsymbol{\alpha}) + \hat{\mathbf{Y}}(\boldsymbol{\alpha})\, \mathbf{Q}\, \hat{\mathbf{a}}(\boldsymbol{\alpha})\|^2$. It will be shown later that under appropriate conditions $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ in (2.8) converges almost surely to $\mathbf{a}(\boldsymbol{\alpha})$ as n tends to infinity. In other words, $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ is a strongly consistent estimator of $\mathbf{a}(\boldsymbol{\alpha})$. This strong consistency can also been shown to be *uniform* in $\boldsymbol{\alpha}$ so that many properties of $\mathbf{a}(\boldsymbol{\alpha})$, as a deterministic function of $\boldsymbol{\alpha}$, are retained by its estimator $\hat{\mathbf{a}}(\boldsymbol{\alpha})$.

2.3 Relation to the CM Method

It should be pointed out that the PF method presented above is an extension of a frequency estimation procedure recently proposed by He and Kedem (1990) and by Yakowitz (1991). A somewhat more systematic and rigorous treatment of this procedure and its statistical properties can be found in Li and Kedem (1991), and Li, Kedem, and Yakowitz (1991), where the procedure was referred to as CM method (or contraction mapping method). In essence, the CM method deals with the case of a *single* sinusoid (at a time) in ambient noise, whereas the PF method extends it to the general case of *multiple* sinusoids. Indeed, for $q = 1$, we note that (A1) becomes

$$\alpha = -2\, r_1^{\epsilon}(\alpha)/r_0^{\epsilon}(\alpha) \tag{2.9}$$

and the LS estimator in (2.8) reduces to

$$\hat{a}(\alpha) = -\sum_{t=3}^{n} \hat{y}_{t-1}(\alpha)\,\{\hat{y}_t(\alpha) + \hat{y}_{t-2}(\alpha)\} / \sum_{t=3}^{n} \hat{y}_{t-1}^2(\alpha).$$

If we reparametrize the filter by setting $\vartheta := -\alpha/2$, then (2.9) can be written as

$$\vartheta = \rho_1^{\epsilon}(\vartheta)$$

where $\rho_1^{\epsilon}(\vartheta)$ stands for the first-order autocorrelation of $\{\epsilon_t(\alpha)\}$. This relation is readily recognized as being the fundamental property required by the CM method for the parametrization of the filter. Moreover, it can be shown (Li and Kedem, 1991) that under appropriate conditions $\hat{\rho}(\vartheta) := -\hat{a}(\alpha)/2$ is a consistent estimator of the first-order autocorrelation of $\{y_t(\alpha)\}$. With this estimator, the fixed-point iteration in (2.6) becomes

$$\hat{\vartheta}_m = \hat{\rho}(\hat{\vartheta}_{m-1}) \qquad (m = 1, 2, \ldots),$$

which coincides with the iteration of the CM method that produces a sequence $\{\hat{\vartheta}_m\}$ for estimating $\vartheta^* := -\alpha^*/2 = \cos\omega_1$.

Statistical properties of the CM method have recently been studied by Li and Kedem (1991) and Li, Kedem, and Yakowitz (1991). In these works, it was proved that under appropriate conditions the CM method provides a strongly consistent estimator of ω_1, and that the estimator is asymptotically normal with a variance inversely related to the signal-to-noise ratio of the data. In the next section, we shall analyze the PF method along the same lines as in these works in order to establish the strong consistency and asymptotic normality of the PF estimator $\hat{\boldsymbol{\alpha}}$.

2.4 Nonsingularity of Autocovariance Matrices

To end this section, let us consider the conditions under which $\mathbf{R}_x(\boldsymbol{\alpha})$ and $\mathbf{R}_\epsilon(\boldsymbol{\alpha})$ are nonsingular so that the inverse matrices in (2.3) and (2.4) are well-defined in the derivation of the PF method. For this purpose, we first extend $\beta_k(\boldsymbol{\alpha})$ and ω_k for $k = q+1, \ldots, 2q$ by defining

$$\beta_{2q-k+1}(\boldsymbol{\alpha}) := \beta_k(\boldsymbol{\alpha}) \quad \text{and} \quad \omega_{2q-k+1} := -\omega_k \qquad (k = 1, \ldots, q).$$

In so doing, we can write the autocovariance function $r_\tau^x(\boldsymbol{\alpha})$ as

$$r_\tau^x(\boldsymbol{\alpha}) = \sum_{k=1}^{q} \tfrac{1}{2}\beta_k^2(\boldsymbol{\alpha}) \cos(\omega_k \tau) = \sum_{k=1}^{2q} \tfrac{1}{4}\beta_k^2(\boldsymbol{\alpha})\, z_k^\tau$$

with $z_k := \exp(i\omega_k)$. It is not difficult to verify from this expression that $\mathbf{R}_x(\boldsymbol{\alpha})$ can be decomposed as

$$\mathbf{R}_x(\boldsymbol{\alpha}) = \mathbf{Q}^T \mathbf{S} \mathbf{P} \mathbf{S}^H \mathbf{Q}, \tag{2.10}$$

where the superscript H denotes the Hermitian transpose. In this decomposition, $\mathbf{P}$ is a $2q$-by-$2q$ diagonal matrix of the form

$$\mathbf{P} := \tfrac{1}{4}\ \mathrm{diag}\{\beta_1^2(\boldsymbol{\alpha}), \ldots, \beta_{2q}^2(\boldsymbol{\alpha})\}$$

and $\mathbf{S}$ is a $(2q-1)$-by-$2q$ Vandermonde matrix as specified by

$$\mathbf{S} := [\mathbf{s}_1, \cdots, \mathbf{s}_{2q}] \tag{2.11}$$

with $\mathbf{s}_k := [1, z_k, \cdots, z_k^{2q-2}]^T$. Since $\mathbf{Q}$ is of full column rank q and $\mathbf{S}^H$ of full column rank $2q-1$, the decomposition in (2.10) indicates that $\mathbf{R}_x(\boldsymbol{\alpha})$ will have full rank q if $\mathbf{P}$ is nonsingular. It is easy to see that the nonsingularity of $\mathbf{P}$ is guaranteed if $\beta_k(\boldsymbol{\alpha}) > 0$, or, equivalently,

(A2) $\qquad H(\omega_k; \boldsymbol{\alpha}) \neq 0 \qquad$ for all $k = 1, \ldots, q$ and $\boldsymbol{\alpha} \in \mathcal{A}$.

That is, the filter should pass *all* frequencies of the signal. Therefore, under (A2), $\mathbf{R}_x(\boldsymbol{\alpha})$ is nonsingular and hence the derivation of (2.3) is valid. Moreover, the nonsingularity of $\mathbf{R}_\epsilon(\boldsymbol{\alpha})$ is guaranteed if $r_0^\epsilon(\boldsymbol{\alpha}) = \mathrm{var}\{\epsilon_t(\boldsymbol{\alpha})\} > 0$ (see, e.g., Brockwell and Davis, 1987, Proposition 5.1.1). By the continuity of $H(\omega; \boldsymbol{\alpha})$ in ω, $r_0^\epsilon(\boldsymbol{\alpha}) = 0$ if and only if $H(\omega; \boldsymbol{\alpha}) \equiv 0$ for all $\omega \in (-\pi, \pi]$. Hence $\mathbf{R}_\epsilon(\boldsymbol{\alpha})$ is also nonsingular under (A2).

3 Statistical Properties of the PF Estimator

To investigate statistical properties of the PF method presented in the preceding section, we shall answer the following questions:

i) Under what conditions does the random mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ have a fixed-point $\hat{\boldsymbol{\alpha}}$?

ii) Under what conditions does the fixed-point iteration in (2.6) converge to the fixed-point $\hat{\boldsymbol{\alpha}}$? and

iii) What limit and limiting distribution does the PF estimator $\hat{\boldsymbol{\alpha}}$ possess as the sample size n tends to infinity?

This section provides a set of sufficient conditions under which a unique fixed-point $\hat{\boldsymbol{\alpha}}$ exists and can be found in the vicinity of $\boldsymbol{\alpha}^*$ by the fixed-point iteration almost surely,. Under these conditions, $\hat{\boldsymbol{\alpha}}$ is also shown to be strongly consistent and asymptotically normal for estimating $\boldsymbol{\alpha}^*$. These results are formulated in terms of the LS estimator $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ defined by (2.8).

3.1 Existence and Convergence

Suppose that the filter also satisfies the following regularity conditions:

(A3) There exist constants $c_j > 0$ with $\sum_{j=0}^{\infty} j c_j < \infty$ such that $|h_j(\boldsymbol{\alpha})| \leq c_j$ for all $j = 0, 1, \ldots$ and $\boldsymbol{\alpha} \in \mathcal{A}$.

(A4) The $h_j(\boldsymbol{\alpha})$ are continuously differentiable in $\mathcal{A}$ and there exist constants $d_{jk} > 0$ with $\sum_{j=0}^{\infty} j d_{jk} < \infty$ such that $|\partial h_j(\boldsymbol{\alpha})/\partial \alpha_k| \leq d_{jk}$ for all $k = 1, \ldots, q$, $j = 0, 1, \ldots$, and $\boldsymbol{\alpha} \in \mathcal{A}$.

Under these conditions, we shall first show that the random mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ is *uniformly* consistent for estimating $\mathbf{a}(\boldsymbol{\alpha})$ up to the first derivative, as summarized by the following lemma.

Lemma 3.1 *Suppose that* (A1)–(A4) *are satisfied. Then,* $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ *and* $\mathbf{a}(\boldsymbol{\alpha})$ *are continuously differentiable, and, as* $n \to \infty$,

$$\hat{\mathbf{a}}(\boldsymbol{\alpha}) \overset{a.s.}{\to} \mathbf{a}(\boldsymbol{\alpha}) \quad \textit{and} \quad \hat{\mathbf{a}}'(\boldsymbol{\alpha}) \overset{a.s.}{\to} \mathbf{a}'(\boldsymbol{\alpha})$$

unifromly in $\boldsymbol{\alpha} \in \mathcal{A}$, *where* $\hat{\mathbf{a}}'(\boldsymbol{\alpha})$ *and* $\mathbf{a}'(\boldsymbol{\alpha})$ *are Jacobian matrices of* $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ *and* $\mathbf{a}(\boldsymbol{\alpha})$, *respectively.*

PROOF. The proof resorts to some basic results of Li and Kedem (1991), concerning the uniform strong consistency of sample autocovariances and their derivatives of the filtered data. In particular, according to Lemma 5.2 and Remark 5.1 of Li and Kedem (1991), (A3) guarantees that

$$n^{-1}\hat{\mathbf{Y}}^T(\boldsymbol{\alpha})\,\hat{\mathbf{Y}}(\boldsymbol{\alpha}) \stackrel{a.s.}{\to} \bar{\mathbf{R}}_y(\boldsymbol{\alpha}) \quad \text{and} \quad n^{-1}\hat{\mathbf{Y}}^T(\boldsymbol{\alpha})\,\hat{\mathbf{y}}(\boldsymbol{\alpha}) \stackrel{a.s.}{\to} \bar{\mathbf{r}}_y(\boldsymbol{\alpha}) + \bar{\mathbf{r}}_y^B(\boldsymbol{\alpha})$$

uniformly in $\boldsymbol{\alpha} \in \mathcal{A}$ as $n \to \infty$. Since $\mathbf{R}_y(\boldsymbol{\alpha})$ is nonsingular under (A2), it follows that $\hat{\mathbf{Y}}^\dagger(\boldsymbol{\alpha})$ is well-defined almost surely for sufficiently large n and

$$\hat{\mathbf{a}}(\boldsymbol{\alpha}) \stackrel{a.s.}{\to} -\mathbf{R}_y^{-1}(\boldsymbol{\alpha})\,\mathbf{r}_y(\boldsymbol{\alpha}) = \mathbf{a}(\boldsymbol{\alpha})$$

uniformly in $\boldsymbol{\alpha} \in \mathcal{A}$. Applying Lemma 5.2 and Remark 5.1 of Li and Kedem (1991) again shows that $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ and $\mathbf{a}(\boldsymbol{\alpha})$ are continuously differentiable under (A4) and $\hat{\mathbf{a}}'(\boldsymbol{\alpha}) \stackrel{a.s.}{\to} \mathbf{a}'(\boldsymbol{\alpha})$ uniformly in $\boldsymbol{\alpha} \in \mathcal{A}$. ◇

In the sequel, we shall also make use of the following lemma.

Lemma 3.2 *Let ϱ be the spectral radius of the matrix $\mathbf{C}(\boldsymbol{\alpha}^*)$, where*

$$\mathbf{C}(\boldsymbol{\alpha}) := \mathbf{R}_y^{-1}(\boldsymbol{\alpha})\,\mathbf{R}_\epsilon(\boldsymbol{\alpha}). \tag{3.1}$$

Then (A2) *implies $\varrho < 1$.*

PROOF. It is easy to verify from (3.1) that $\mathbf{C}(\boldsymbol{\alpha})$ can also be written as

$$\mathbf{C}(\boldsymbol{\alpha}) = \{\mathbf{I} + \boldsymbol{\Gamma}(\boldsymbol{\alpha})\}^{-1} \quad \text{with} \quad \boldsymbol{\Gamma}(\boldsymbol{\alpha}) := \mathbf{R}_\epsilon^{-1}(\boldsymbol{\alpha})\,\mathbf{R}_x(\boldsymbol{\alpha}). \tag{3.2}$$

Let μ_j and $\mathbf{p}_j$, $(j = 1, \ldots, q)$, be the eigenvalues and corresponding eigenvectors of $\boldsymbol{\Gamma}(\boldsymbol{\alpha}^*)$, then $1/(1 + \mu_j)$ are eigenvalues of $\mathbf{C}(\boldsymbol{\alpha}^*)$, associated with eigenvectors $\mathbf{p}_j$. By definition (Ortega and Rheinboldt, 1970, p. 43), $\varrho = \max\{1/|1 + \mu_j|\}$. Therefore, $\varrho < 1$ if $|1 + \mu_j| > 1$ for all j. On the other hand, since $\boldsymbol{\Gamma}(\boldsymbol{\alpha}^*)\,\mathbf{p}_j = \mu_j \mathbf{p}_j$, it follows from (3.2) that

$$\mathbf{p}_j^H \mathbf{R}_x(\boldsymbol{\alpha}^*)\,\mathbf{p}_j = \mu_j\,\{\mathbf{p}_j^H \mathbf{R}_\epsilon(\boldsymbol{\alpha}^*)\,\mathbf{p}_j\}.$$

Note that $\mathbf{R}_x(\boldsymbol{\alpha}^*)$ and $\mathbf{R}_\epsilon(\boldsymbol{\alpha}^*)$ are positive definite under (A2). Therefore, we obtain $\mathbf{p}_j^H \mathbf{R}_x(\boldsymbol{\alpha}^*)\,\mathbf{p}_j > 0$ and $\mathbf{p}_j^H \mathbf{R}_\epsilon(\boldsymbol{\alpha}^*)\,\mathbf{p}_j > 0$ for all j. This, in turn, yields $\mu_j > 0$ for all j. As a consequence, we obtain $|1+\mu_j| = 1+\mu_j > 1$ for all j and hence $\varrho < 1$. $\diamondsuit$

Based on these lemmas, the existence of the PF estimator $\hat{\boldsymbol{\alpha}}$ as a fixed-point of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ and the convergence of FPI to $\hat{\boldsymbol{\alpha}}$ can be established as follows.

Theorem 3.1 *Under* (A1)–(A4), *the following assertions hold almost surely, provided that n is sufficiently large.*

a) There exists a neighborhood $S_\Delta(\boldsymbol{\alpha}^*) := \{\boldsymbol{\alpha} : \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\| < \Delta\}$ of $\boldsymbol{\alpha}^*$, with Δ being independent of n, in which the random mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ has a unique fixed-point $\hat{\boldsymbol{\alpha}}$.

b) The sequence $\{\hat{\boldsymbol{\alpha}}_m\}$ produced by (2.6) *converges to $\hat{\boldsymbol{\alpha}}$ as $m \to \infty$ if $\hat{\boldsymbol{\alpha}}_0 \in S_\delta(\hat{\boldsymbol{\alpha}})$, where $S_\delta(\hat{\boldsymbol{\alpha}}) \subset S_\Delta(\boldsymbol{\alpha}^*)$ is a neighborhood of $\hat{\boldsymbol{\alpha}}$, with δ being independent of n.*

PROOF. Notice that (2.2) yields $\mathbf{a}(\boldsymbol{\alpha}) - \mathbf{a}(\boldsymbol{\alpha}^*) = -\mathbf{R}_y^{-1}(\boldsymbol{\alpha})\,\{\mathbf{r}_\epsilon(\boldsymbol{\alpha}) + \mathbf{R}_\epsilon(\boldsymbol{\alpha})\,\mathbf{a}\}$. Since (A1) is equivalent to $\boldsymbol{\alpha} = -\mathbf{R}_c^{-1}(\boldsymbol{\alpha})\,\mathbf{r}_\epsilon(\boldsymbol{\alpha})$, it is easy to verify that

$$\mathbf{a}(\boldsymbol{\alpha}) - \mathbf{a}(\boldsymbol{\alpha}^*) = \mathbf{R}_y^{-1}(\boldsymbol{\alpha})\,\mathbf{R}_\epsilon(\boldsymbol{\alpha})\,(\boldsymbol{\alpha} - \mathbf{a}) = \mathbf{C}(\boldsymbol{\alpha})\,(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*). \tag{3.3}$$

It follows from the continuity of $\mathbf{C}(\boldsymbol{\alpha})$ that the Jacobian matrix of $\mathbf{a}(\boldsymbol{\alpha})$ at $\boldsymbol{\alpha}^*$ is given by $\mathbf{a}'(\boldsymbol{\alpha}^*) = \mathbf{C}(\boldsymbol{\alpha}^*)$. By Lemma 3.2, the spectral radius of $\mathbf{a}'(\boldsymbol{\alpha}^*)$ is strictly less than 1. Thus (see, e.g., Ortega and Rheinboldt, 1970, Theorem 2.2.8, p. 44), there exits a norm $\|\cdot\|$ such that $\|\mathbf{a}'(\boldsymbol{\alpha}^*)\| < 1$. Furthermore, the continuity of $\mathbf{a}'(\boldsymbol{\alpha})$ and $\mathbf{C}(\boldsymbol{\alpha})$ also guarantees the existence of a constant $0 < c < 1$ and a neighborhood $\bar{S}_\Delta(\boldsymbol{\alpha}^*) := \{\boldsymbol{\alpha} : \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\| \le \Delta\} \subseteq \mathcal{A}$ such that

$$\|\mathbf{a}'(\boldsymbol{\alpha})\| \le c \quad \text{and} \quad \|\mathbf{C}(\boldsymbol{\alpha})\| \le c \tag{3.4}$$

for all $\boldsymbol{\alpha} \in \bar{S}_\Delta(\boldsymbol{\alpha}^*)$. Let $\kappa := (c+1)/2 < 1$, then, for any $0 < \Delta_0 < \Delta$, the uniform convergence of $\hat{\mathbf{a}}'(\boldsymbol{\alpha})$ to $\mathbf{a}'(\boldsymbol{\alpha})$ implies that $\|\hat{\mathbf{a}}'(\boldsymbol{\alpha})\| \le \kappa$ almost surely for all $\boldsymbol{\alpha} \in \bar{S}_{\Delta_0}(\boldsymbol{\alpha}^*)$, provided that n is sufficiently large. On the other hand, using the mean-value theorem (Ortega and Rheinboldt, 1970, p. 71) we can show that almost surely

$$\|\hat{\mathbf{a}}(\boldsymbol{\alpha}_1) - \hat{\mathbf{a}}(\boldsymbol{\alpha}_2)\| \le \kappa\,\|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\| \tag{3.5}$$

for all $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2 \in \bar{S}_{\Delta_0}(\boldsymbol{\alpha}^*)$, that is, the random mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ is *contractive* on $\bar{S}_{\Delta_0}(\boldsymbol{\alpha}^*)$ almost surely, provided that n is sufficiently large. In addition, the convergence of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ to $\mathbf{a}(\boldsymbol{\alpha})$, as shown in Lemma 3.1, guarantees that

$$\|\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) - \boldsymbol{\alpha}^*\| = \|\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) - \mathbf{a}(\boldsymbol{\alpha}^*)\| \leq (1 - \kappa)\Delta_0$$

almost surely for sufficiently large n. Therefore, according to Theorem 5.2.3 of Stoer and Bulirsch (1980), the mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ has a unique fixed-point $\hat{\boldsymbol{\alpha}}$ on $\bar{S}_{\Delta_0}(\boldsymbol{\alpha}^*)$. Since Δ_0 is arbitrary, $\hat{\boldsymbol{\alpha}}$ must be a unique fixed-point of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ in the *interior* of $S_\Delta(\boldsymbol{\alpha}^*)$. Part a) of the theorem is thus proved. To show Part b), we note that for any neighborhood $S_\delta(\hat{\boldsymbol{\alpha}})$ of the fixed-point $\hat{\boldsymbol{\alpha}}$, if $S_\delta(\hat{\boldsymbol{\alpha}}) \subset S_\Delta(\boldsymbol{\alpha}^*)$, then (3.5) remains valid almost surely for all $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2 \in S_\delta(\hat{\boldsymbol{\alpha}})$, that is, $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ is almost surely a *contraction mapping* on $S_\delta(\hat{\boldsymbol{\alpha}})$, provided that n is sufficiently large. By Theorem 5.2.2 of Stoer and Bulirsch (1980), the sequence $\{\hat{\boldsymbol{\alpha}}_m\}$ produced by (2.6) converges to $\hat{\boldsymbol{\alpha}}$ as $m \to \infty$ almost surely, if n is sufficiently large and the initial value $\hat{\boldsymbol{\alpha}}_0$ is chosen in $S_\delta(\hat{\boldsymbol{\alpha}})$. $\diamondsuit$

REMARK. Theory of numerical analysis (see, e.g., Stoer and Bulirsch, 1980) tells us that the spectral radius ϱ of $\mathbf{C}(\boldsymbol{\alpha}^*)$ is crucial to the rate of convergence of FPI. Indeed, the smaller the spectral radius ϱ is, the faster is the convergence of FPI to $\hat{\boldsymbol{\alpha}}$. As seen in the proof of Lemma 3.2, $\varrho = 1/(1 + \mu_{\min})$, where $\mu_{\min} := \min\{\mu_j\}$. Therefore, to accelerate the convergence of FPI, $\mu_{\min}$ should be made as large as possible. Notice that

$$\mu_{\min} = \min_{\mathbf{p} \neq \mathbf{0}} \frac{\mathbf{p}^H \mathbf{R}_x(\boldsymbol{\alpha}^*)\,\mathbf{p}}{\mathbf{p}^H \mathbf{R}_\epsilon(\boldsymbol{\alpha}^*)\,\mathbf{p}}.$$

In the case of $q = 1$, $\mu_{\min}$ recudes to $r_0^x(\boldsymbol{\alpha}^*)/r_0^\epsilon(\boldsymbol{\alpha}^*)$, which is readily recognized to be the signal-to-noise ratio of the filtered process $\{y_t(\boldsymbol{\alpha})\}$ with the desired filter parameter $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$. For $q > 1$, $\mu_{\min}$ can be regarded as a generalized indicator of the amount of signal relative to the amount of noise in the filtered process $\{y_t(\boldsymbol{\alpha}^*)\}$.

3.2 Strong Consistency and Asymptotic Normality

Suppose that $\hat{\boldsymbol{\alpha}}$ is the fixed-point of the random mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ in the vicinity of $\boldsymbol{\alpha}^*$. In the following, we shall investigate asymptotic properties of $\hat{\boldsymbol{\alpha}}$ as the sample size n tends to infinity. The following theorem claims the strong consistency of the PF estimator $\hat{\boldsymbol{\alpha}}$ for estimating $\boldsymbol{\alpha}^*$.

Theorem 3.2 *Suppose that* (A1)–(A4) *are satisfied, and let* $\hat{\boldsymbol{\alpha}}$ *be the unique fixed-point of* $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ *in* $S_\Delta(\boldsymbol{\alpha}^*)$, *as given by Theorem* 3.1. *Then* $\hat{\boldsymbol{\alpha}}$ *converges to* $\boldsymbol{\alpha}^*$ *almost surely as* $n \to \infty$.

PROOF. Since $\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}}) = \hat{\boldsymbol{\alpha}}$ and $\mathbf{a}(\boldsymbol{\alpha}^*) = \boldsymbol{\alpha}^*$, it follows from (3.3) that

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^* = \delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}}) + \mathbf{a}(\hat{\boldsymbol{\alpha}}) - \mathbf{a}(\boldsymbol{\alpha}^*) = \delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}}) + \mathbf{C}(\hat{\boldsymbol{\alpha}})\,(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) \tag{3.6}$$

where $\delta\hat{\mathbf{a}}(\boldsymbol{\alpha}) := \hat{\mathbf{a}}(\boldsymbol{\alpha}) - \mathbf{a}(\boldsymbol{\alpha})$. By (3.4), we have $\|\mathbf{C}(\boldsymbol{\alpha})\| \le c < 1$ for any $\boldsymbol{\alpha} \in S_\Delta(\boldsymbol{\alpha}^*)$. It turns out that

$$\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\| \le \|\delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}})\| + \|\mathbf{C}(\hat{\boldsymbol{\alpha}})\|\,\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\| \le \|\delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}})\| + c\,\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|$$

and hence

$$0 \le (1-c)\,\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\| \le \|\delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}})\| \tag{3.7}$$

almost surely for large n. The uniform convergence of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ to $\mathbf{a}(\boldsymbol{\alpha})$, as shown in Lemma 3.1, guarantees that $\|\delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}})\| \xrightarrow{a.s.} 0$ as $n \to \infty$, which, together with (3.7), proves the assertion. ◇

The asymptotic normality can also be established for the PF estimator $\hat{\boldsymbol{\alpha}}$. To this end, we first need the following lemma, concerning the asymptotic normality of $\hat{\mathbf{a}}(\boldsymbol{\alpha}^*)$.

Lemma 3.3 *Suppose that* (A1)–(A3) *are satisfied. If in addition* $E(\xi_t^4) < \infty$, *then, as* $n \to \infty$, $\sqrt{n}\,\delta\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) = \sqrt{n}\,\{\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) - \mathbf{a}\}$ *converges in distribution to a normal random vector with mean zero and covariance matrix*

$$\mathbf{V} := \mathbf{R}_y^{-1}(\boldsymbol{\alpha}^*)\,\mathbf{Q}^T\mathbf{W}(\boldsymbol{\alpha}^*)\,\mathbf{Q}\,\mathbf{R}_y^{-1}(\boldsymbol{\alpha}^*)$$

where $\mathbf{W}(\boldsymbol{\alpha}^*) := [w_{ij}(\boldsymbol{\alpha}^*)]$ *and*

$$w_{ij}(\boldsymbol{\alpha}^*) := 4 \sum_{\tau=0}^{\infty} \left\{ \sum_{k=0}^{2q} a_k \, r^c_{\tau+i-k}(\boldsymbol{\alpha}^*) \right\} \left\{ \sum_{k=0}^{2q} a_k \, r^c_{\tau+j-k}(\boldsymbol{\alpha}^*) \right\}. \tag{3.8}$$

for $i, j = 1, \ldots, 2q-1$.

PROOF. The proof utilizes some basic results of Li, Kedem, and Yakowitz (1991), concerning the asymptotic normality of sample autocovariances of the filtered data. First of all, under (A1), it is easy to see that (2.8) yields

$$\delta\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) = -\{(\mathbf{Q}^T\hat{\mathbf{Y}}^T\hat{\mathbf{Y}}\mathbf{Q})^{-1}\mathbf{Q}^T\hat{\mathbf{Y}}^T\hat{\mathbf{y}} + \mathbf{a}\}.$$

Here, as well as in the following, the argument $\boldsymbol{\alpha}^*$ is omitted in data and autocovariance matrices for the sake of brevity. Moreover, by following the same lines as the proof of Theorem 3.1 of Li, Kedem, and Yakowitz (1991), we can write

$$n^{-1}\mathbf{Q}^T\hat{\mathbf{Y}}^T\hat{\mathbf{Y}}\mathbf{Q} = \hat{\mathbf{R}} + o_P(n^{-1/2}) \quad \text{and} \quad n^{-1}\mathbf{Q}^T\hat{\mathbf{Y}}^T\hat{\mathbf{y}} = \hat{\mathbf{r}} + o_P(n^{-1/2}), \tag{3.9}$$

where $\hat{\mathbf{R}} := \mathbf{Q}^T\check{\mathbf{R}}\mathbf{Q}$, $\hat{\mathbf{r}} := \mathbf{Q}^T(\check{\mathbf{r}} + \check{\mathbf{r}}^B) = 2\,\mathbf{Q}^T\check{\mathbf{r}}$, with

$$\check{\mathbf{R}} := \begin{bmatrix} \hat{r}_0 & \cdots & \hat{r}_{-2q+2} \\ \vdots & \ddots & \vdots \\ \hat{r}_{2q-2} & \cdots & \hat{r}_0 \end{bmatrix} \qquad \check{\mathbf{r}} := \begin{bmatrix} \hat{r}_1 \\ \vdots \\ \hat{r}_{2q-1} \end{bmatrix}$$

and

$$\hat{r}_j := n^{-1} \sum_{t=1}^{n-j} \hat{y}_{t+j}(\boldsymbol{\alpha}^*)\, \hat{y}_t(\boldsymbol{\alpha}^*) \qquad (j = 0, 1, \ldots, 2q-1).$$

Therefore, it follows from (3.9) that

$$\begin{aligned} \delta\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) &= -(\hat{\mathbf{R}}^{-1}\hat{\mathbf{r}} + \mathbf{a}) + o_P(n^{-1/2}) \\ &= -\hat{\mathbf{R}}^{-1}(\hat{\mathbf{r}} + \hat{\mathbf{R}}\mathbf{a}) + o_P(n^{-1/2}). \end{aligned} \tag{3.10}$$

Let $r_j := r^y_j(\boldsymbol{\alpha}^*)$ for brevity. Then, according to Theorem 3.1 and Remark 3.3 of Li, Kedem, and Yakowitz (1991), $\sqrt{n}\,\{\hat{r}_j - r_j\}, (j = 0, \ldots, 2q-1)$ are asymptotically jointly

normal with mean zero and covariance matrix $\mathbf{V}_r := [v_{ij}]$, where

$$\begin{aligned} v_{ij} &:= 2\sum_{k=1}^{q} \beta_k^2(\boldsymbol{\alpha}^*)\,\cos(i\omega_k)\cos(j\omega_k) \sum_{\tau=-\infty}^{\infty} r_\tau^\epsilon(\boldsymbol{\alpha}^*)\,\cos(\tau\omega_k) \\ &\quad + (\mu-3)\, r_i^\epsilon(\boldsymbol{\alpha}^*)\, r_j^\epsilon(\boldsymbol{\alpha}^*) \\ &\quad + \sum_{\tau=-\infty}^{\infty} \{r_\tau^\epsilon(\boldsymbol{\alpha}^*)\, r_{\tau+i-j}^\epsilon(\boldsymbol{\alpha}^*) + r_{\tau+i}^\epsilon(\boldsymbol{\alpha}^*)\, r_{\tau-j}^\epsilon(\boldsymbol{\alpha}^*)\} \end{aligned} \tag{3.11}$$

for $i, j = 0, 1, \ldots, 2q-1$, and $\mu := E\{\epsilon_t^4(\boldsymbol{\alpha}^*)\}/E\{\epsilon_t^2(\boldsymbol{\alpha}^*)\} < \infty$. On the other hand, the vector $\hat{\mathbf{r}} + \hat{\mathbf{R}}\mathbf{a}$ in (3.10) can be regarded as the value of some function $\mathbf{f}(\zeta_0, \ldots, \zeta_{2q-1})$ at $\zeta_j = \hat{r}_j$, that is, $\hat{\mathbf{r}} + \hat{\mathbf{R}}\mathbf{a} = \mathbf{f}(\hat{r}_0, \ldots, \hat{r}_{2q-1})$. For this function, it is also true, by (2.1), that

$$\begin{aligned} \mathbf{f}(r_0, \ldots, r_{2q-1}) &= \mathbf{r}_y(\boldsymbol{\alpha}^*) + \mathbf{R}_y(\boldsymbol{\alpha}^*)\,\mathbf{a} \\ &= \mathbf{r}_y(\boldsymbol{\alpha}^*) + \mathbf{R}_y(\boldsymbol{\alpha}^*)\,\mathbf{a}(\boldsymbol{\alpha}^*) \;=\; \mathbf{0}. \end{aligned}$$

Therefore, invoking Proposition 6.4.3 of Brockwell and Davis (1987) proves that

$$\sqrt{n}\,(\hat{\mathbf{r}} + \hat{\mathbf{R}}\mathbf{a}) = \sqrt{n}\,\{\mathbf{f}(\hat{r}_0, \ldots, \hat{r}_{2q-1}) - \mathbf{f}(r_0, \ldots, r_{2q-1})\}$$

converges in distribution to a normal random vector with mean zero and covariance matrix $\mathbf{V}_f := \mathbf{F}\mathbf{V}_r\mathbf{F}^T$, where $\mathbf{F}$ is the Jacobian matrix of $\mathbf{f}$ evaluated at $(r_0, \ldots, r_{2q-1})$. It is easy to verify that

$$\hat{\mathbf{r}} + \hat{\mathbf{R}}\mathbf{a} = \mathbf{Q}^T(\check{\mathbf{r}} + \check{\mathbf{r}}^B + \check{\mathbf{R}}\mathbf{Q}\mathbf{a}) = \mathbf{Q}^T[\check{\mathbf{r}} \,\vdots\, \check{\mathbf{R}} \,\vdots\, \check{\mathbf{r}}^B]\,\bar{\mathbf{a}},$$

where $\bar{\mathbf{a}} := [a_0, a_1, \cdots, a_{2q}]^T$. Simple algebra shows $\mathbf{F} = [\mathbf{f}_0, \cdots, \mathbf{f}_{2q-1}]$, where

$$\mathbf{f}_0 = \mathbf{Q}^T \begin{bmatrix} a_1 \\ \vdots \\ a_{2q-1} \end{bmatrix} \quad \text{and} \quad \mathbf{f}_j = \mathbf{Q}^T \begin{bmatrix} a_{1-j} + a_{1+j} \\ \vdots \\ a_{2q-1-j} + a_{2q-1+j} \end{bmatrix}$$

for $j = 1, \ldots, 2q-1$, with $a_k := 0$ for $k < 0$ and $k > 2q$. Upon noting that v_{ij} given by (3.11) are symmetric in the sense that $v_{-i,j} = v_{i,-j} = v_{-i,-j} = v_{ij}$, we can rewrite $\mathbf{V}_f$ as $\mathbf{V}_f = \mathbf{Q}^T\mathbf{B}\mathbf{Q}$, where $\mathbf{B} := [b_{ij}]$, $(i, j = 1, \ldots, 2q-1)$, with

$$b_{ij} := \sum_{k,l=0}^{2q} v_{i-k,j-l}\; a_k\, a_l\,.$$

As can be seen from (3.11), there are three groups in the expression of v_{ij}. The first group involves the sinusoidal terms, all of which are cancelled out in the expression of b_{ij}, because $\sum_{l=0}^{2q} a_l z_k^{-l} = 0$ for $k = 1, \ldots, q$ and hence

$$\sum_{l=0}^{2q} a_l \cos\{\omega_k(j-l)\} = 0$$

for $j = 1, \ldots, 2q-1$ and $k = 1, \ldots, q$. The second group in (3.11) consists of $v'_{ij} := (\mu - 3)\, r_i^{\epsilon}(\boldsymbol{\alpha}^*)\, r_j^{\epsilon}(\boldsymbol{\alpha}^*)$. It is easy to see that the corresponding term in $\mathbf{V}_f$ can be written as $(\mu - 3)\,\mathbf{U}\mathbf{U}^T$, where

$$\begin{aligned} \mathbf{U} &:= \mathbf{Q}^T[\bar{\mathbf{r}}_{\epsilon}(\boldsymbol{\alpha}^*) \,\vdots\, \bar{\mathbf{R}}_{\epsilon}(\boldsymbol{\alpha}^*) \,\vdots\, \bar{\mathbf{r}}_{\epsilon}^B(\boldsymbol{\alpha}^*)]\,\bar{\mathbf{a}} \\ &= \mathbf{r}_{\epsilon}(\boldsymbol{\alpha}^*) + \mathbf{R}_{\epsilon}(\boldsymbol{\alpha}^*)\,\mathbf{a} \;=\; \mathbf{r}_{\epsilon}(\boldsymbol{\alpha}^*) + \mathbf{R}_{\epsilon}(\boldsymbol{\alpha}^*)\,\boldsymbol{\alpha}^*. \end{aligned}$$

Under (A1), $\mathbf{U} = \mathbf{0}$ and hence that term in $\mathbf{V}_f$ vanishes. Combining these results, we obtain $\mathbf{V}_f = \mathbf{Q}^T\mathbf{B}_0\,\mathbf{Q}$ where $\mathbf{B}_0 := [b'_{ij}]$ with

$$b'_{ij} := \sum_{\tau=-\infty}^{\infty} \sum_{k,\,l=0}^{2q} a_k\, a_l\, r_{\tau+i-k}^{\epsilon}(\boldsymbol{\alpha}^*)\{r_{\tau+j-l}^{\epsilon}(\boldsymbol{\alpha}^*) + r_{\tau-j+l}^{\epsilon}(\boldsymbol{\alpha}^*)\}.$$

Furthermore, it is not difficult to verify that $\mathbf{B}_0$ can be written compactly as

$$\mathbf{B}_0 = \sum_{\tau=-\infty}^{\infty} \mathbf{A}^T\mathbf{r}_{\tau}(\boldsymbol{\alpha}^*)\,\mathbf{r}_{\tau}^T(\boldsymbol{\alpha}^*)\,(\mathbf{A} + \tilde{\mathbf{I}}\mathbf{A}),$$

where $\mathbf{A}$ is a $(4q-1)$-by-$(2q-1)$ matrix of the form

$$\mathbf{A} := \begin{bmatrix} a_{2q} & & 0 \\ \vdots & \ddots & \\ a_0 & & a_{2q} \\ & \ddots & \vdots \\ 0 & & a_0 \end{bmatrix}$$

and $\mathbf{r}_{\tau}(\boldsymbol{\alpha}^*) := [r_{\tau-2q+1}^{\epsilon}(\boldsymbol{\alpha}^*), \cdots, r_{\tau+2q-1}^{\epsilon}(\boldsymbol{\alpha}^*)]^T$. Since $\tilde{\mathbf{I}}\mathbf{A} = \mathbf{A}\tilde{\mathbf{I}}$ and $\tilde{\mathbf{I}}\mathbf{Q} = \mathbf{Q}$, we obtain $\tilde{\mathbf{I}}\mathbf{A}\mathbf{Q} = \mathbf{A}\mathbf{Q}$. Therefore,

$$\mathbf{V}_f = 2 \sum_{\tau=-\infty}^{\infty} \mathbf{Q}^T\mathbf{A}^T\mathbf{r}_{\tau}(\boldsymbol{\alpha}^*)\,\mathbf{r}_{\tau}^T(\boldsymbol{\alpha}^*)\,\mathbf{A}\mathbf{Q}.$$

This expression can be further simplified as

$$\mathbf{V}_f = 4 \sum_{\tau=0}^{\infty} \mathbf{Q}^T \mathbf{A}^T \mathbf{r}_\tau(\boldsymbol{\alpha}^*)\, \mathbf{r}_\tau^T(\boldsymbol{\alpha}^*)\, \mathbf{A}\mathbf{Q} = \mathbf{Q}^T \mathbf{W}(\boldsymbol{\alpha}^*)\, \mathbf{Q},$$

upon noting that $\mathbf{r}_{-\tau}(\boldsymbol{\alpha}) = \tilde{\mathbf{I}}\, \mathbf{r}_\tau(\boldsymbol{\alpha})$, $\mathbf{Q}^T \mathbf{A}^T \mathbf{r}_0(\boldsymbol{\alpha}^*) = \mathbf{U} = \mathbf{0}$, and

$$\mathbf{W}(\boldsymbol{\alpha}^*) = 4 \sum_{\tau=0}^{\infty} \mathbf{A}^T \mathbf{r}_\tau(\boldsymbol{\alpha}^*)\, \mathbf{r}_\tau^T(\boldsymbol{\alpha}^*)\, \mathbf{A}.$$

Finally, since $\hat{\mathbf{R}} \xrightarrow{a.s.} \mathbf{R}_y(\boldsymbol{\alpha}^*)$, by Slutsky's theorem (Lehmann, 1982, Lemma 4.1, pp. 432–433) we obtain from (3.10) that $\sqrt{n}\, \delta\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) = -\hat{\mathbf{R}}^{-1} \sqrt{n}\, (\hat{\mathbf{r}} + \hat{\mathbf{R}}\mathbf{a}) + o_P(1)$ converges in distribution to $\mathrm{N}(\mathbf{0}, \mathbf{V})$. ◇

With the aid of this lemma, we are now able to show the asymptotic normality of the PF estimator $\hat{\boldsymbol{\alpha}}$.

Theorem 3.3 *Suppose that the conditions in Theorem* 3.2 *are satisfied and that* $E(\xi_t^4) < \infty$. *Then, as* $n \to \infty$, $\sqrt{n}\, (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)$ *converges in distribution to a normal random vector with mean zero and covariance matrix*

$$\mathbf{V}_\alpha = \mathbf{R}_x^{-1}(\boldsymbol{\alpha}^*)\, \mathbf{Q}^T \mathbf{W}(\boldsymbol{\alpha}^*)\, \mathbf{Q}\, \mathbf{R}_x^{-1}(\boldsymbol{\alpha}^*) \tag{3.12}$$

where $\mathbf{W}(\boldsymbol{\alpha}^*)$ *is defined in Lemma* 3.3.

PROOF. It follows from (3.6) that $\{\mathbf{I} - \mathbf{C}(\hat{\boldsymbol{\alpha}})\}\, (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) = \delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}})$. By the mean-value theorem, $\delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}})$ can be written as

$$\delta\hat{\mathbf{a}}(\hat{\boldsymbol{\alpha}}) = \delta\hat{\mathbf{a}}(\boldsymbol{\alpha}^*) + \left\{ \int_0^1 \delta\hat{\mathbf{a}}'(\boldsymbol{\alpha}^* + \lambda(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*))\, d\lambda \right\} (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)$$

where $\delta\hat{\mathbf{a}}'(\boldsymbol{\alpha})$ is the Jacobian matrix of $\delta\hat{\mathbf{a}}(\boldsymbol{\alpha})$. Since $\delta\hat{\mathbf{a}}'(\boldsymbol{\alpha}) = \hat{\mathbf{a}}'(\boldsymbol{\alpha}) - \mathbf{a}'(\boldsymbol{\alpha}) \xrightarrow{a.s.} \mathbf{0}$ uniformly in $\boldsymbol{\alpha} \in \mathcal{A}$, as guaranteed by Lemma 3.1, we have

$$\int_0^1 \delta\hat{\mathbf{a}}'(\boldsymbol{\alpha}^* + \lambda(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*))\, d\lambda \xrightarrow{a.s.} \mathbf{0}.$$

In addition, the consistency of $\hat{\boldsymbol{\alpha}}$, together with the continuity of $\mathbf{C}(\boldsymbol{\alpha})$, implies that $\mathbf{C}(\hat{\boldsymbol{\alpha}}) \xrightarrow{a.s.} \mathbf{C}(\boldsymbol{\alpha}^*)$. Therefore, by Slutsky's theorem, $\sqrt{n}\, (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)$ has the same asymptotic

distribution as $\sqrt{n}\,\{\mathbf{I}-\mathbf{C}(\boldsymbol{\alpha}^*)\}^{-1}\delta\hat{\mathbf{a}}(\boldsymbol{\alpha}^*)$. Invoking Lemma 3.3 shows that $\sqrt{n}\,(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^*)$ converges in distribution to $\mathrm{N}(\mathbf{0},\mathbf{V}_\alpha)$, where $\mathbf{V}_\alpha := \{\mathbf{I}-\mathbf{C}(\boldsymbol{\alpha}^*)\}^{-1}\mathbf{V}\,\{\mathbf{I}-\mathbf{C}(\boldsymbol{\alpha}^*)\}^{-1}$. Furthermore, using the expression (3.2) of $\mathbf{C}(\boldsymbol{\alpha})$ and applying the matrix-inversion formula, we can write

$$\{\mathbf{I}-\mathbf{C}(\boldsymbol{\alpha})\}^{-1} = \mathbf{I}+\boldsymbol{\Gamma}^{-1}(\boldsymbol{\alpha}) = \boldsymbol{\Gamma}^{-1}(\boldsymbol{\alpha})\,\{\mathbf{I}+\boldsymbol{\Gamma}(\boldsymbol{\alpha})\}.$$

On the other hand, $\mathbf{R}_y(\boldsymbol{\alpha}) = \mathbf{R}_x(\boldsymbol{\alpha})+\mathbf{R}_\epsilon(\boldsymbol{\alpha}) = \mathbf{R}_\epsilon(\boldsymbol{\alpha})\,\{\mathbf{I}+\boldsymbol{\Gamma}(\boldsymbol{\alpha})\}$. Therefore, we obtain $\{\mathbf{I}-\mathbf{C}(\boldsymbol{\alpha})\}^{-1}\mathbf{R}_y^{-1}(\boldsymbol{\alpha}) = \boldsymbol{\Gamma}^{-1}(\boldsymbol{\alpha})\,\mathbf{R}_\epsilon^{-1}(\boldsymbol{\alpha}) = \mathbf{R}_x^{-1}(\boldsymbol{\alpha})$. Substituting this result in the above expression of $\mathbf{V}_\alpha$ completes the proof. ◇

4 Extension to Complex Sinusoids

A parallel theory of the PF method can be easily established for the case of complex sinusoids in additive noise. In fact, if the signal $\{x_t\}$ is a sum of p complex sinusoids as given by

$$x_t := \sum_{k=1}^{p} \beta_k\, e^{i(\omega_k t+\phi_k)}$$

with $0<\omega_1<\cdots<\omega_p<2\pi$, then it satisfies a pth-order AR autoregressive equation of the form

$$\sum_{j=0}^{p} a_j\, x_{t-j} = 0$$

where the AR parameter vector $\mathbf{a} := [a_1,\cdots,a_p]^T$ is defined by the coefficients of the polynomial

$$\sum_{j=0}^{p} a_j\, z^{p-j} = \prod_{k=1}^{p}(z-z_k)$$

with $z_k := \exp(i\omega_k)$. Given a finite data set $\{y_1,\ldots,y_n\}$ observed from (1.2), one of the widely-used estimators of the AR parameter $\mathbf{a}$ is given by

$$\mathbf{a}_{\mathrm{LS}} := -(\mathbf{Y}^H\mathbf{Y})^{-1}\mathbf{Y}^H\mathbf{y} \tag{4.13}$$

where $\mathbf{Y}$ and $\mathbf{y}$ are redefined by

$$\mathbf{Y} := \begin{bmatrix} y_p & \cdots & y_1 \\ \vdots & & \vdots \\ y_{n-1} & \cdots & y_{n-p} \end{bmatrix} \quad \text{and} \quad \mathbf{y} := \begin{bmatrix} y_{p+1} \\ \vdots \\ y_n \end{bmatrix}.$$

This estimator is known as the forward linear prediction (FLP) (Kay and Marple, 1981) which minimizes the criterion $\|\mathbf{y}+\mathbf{Ya}\|^2$. Other procedures such as the forward-backward linear prediction (FBLP) (Kay and Marple, 1981) are also applicable for estimating the AR parameter $\mathbf{a}$ with only a slight modification of $\mathbf{Y}$ and $\mathbf{y}$ in (4.13).

Introducing a parametric filter $\{h_j(\boldsymbol{\alpha})\}$ indexed by the parameter $\boldsymbol{\alpha} := [\alpha_1, \cdots, \alpha_p]^T$, an estimator $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ of the AR parameter $\mathbf{a}$ can be obtained according to (4.13), with the data matrices replaced by those of the filtered data $\{\hat{y}_t(\boldsymbol{\alpha})\}$ in (2.7). The assumption (A1) retains its form, but $\mathbf{R}_\epsilon(\boldsymbol{\alpha})$ and $\mathbf{r}_\epsilon(\boldsymbol{\alpha})$ are now of the structure

$$\mathbf{R}_\epsilon := \begin{bmatrix} r_0^\epsilon & r_{-1}^\epsilon & \cdots & r_{-p+1}^\epsilon \\ r_1^\epsilon & r_0^\epsilon & \cdots & r_{-p+2}^\epsilon \\ \vdots & \vdots & \ddots & \vdots \\ r_{p-1}^\epsilon & r_{p-2}^\epsilon & \cdots & r_0^\epsilon \end{bmatrix} \quad \mathbf{r}_\epsilon := \begin{bmatrix} r_1^\epsilon \\ r_2^\epsilon \\ \vdots \\ r_p^\epsilon \end{bmatrix} \tag{4.14}$$

with the autocovariances obtained from the filtered noise $\{\epsilon_t(\boldsymbol{\alpha})\}$. In this case, (A1) is readily recognized as being the Yule-Walker equations. In other words, for complex sinusoids, (A1) can be interpreted as the requirement that the filter be parametrized so that the parameter $\boldsymbol{\alpha}$ satisfis the Yule-Walker equation for the filtered noise. With this property being fulfilled, the PF estimator $\hat{\boldsymbol{\alpha}}$ is defined as the fixed-point of the random mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ and can be obtained by FPI in (2.6).

5 The AR Filter

Although its consistency is guaranteed by the asymptotic theory as developed in Section 3, the accuracy of the PF estimator depends on the choice of the parametric filter to be applied to the data. Intuitively, a "good" filter should be bandpass, so that the

sinusoidal signal can be enhanced and the noise be eliminated as much as possible. A useful example of such a filter is the AR filter that will be considered in detail as an illustration of the PF method.

The AR($2q$) filter (or simply the AR filter) is a parametric filter defined recursively by

$$y_t(\boldsymbol{\alpha}) + \theta_1(\boldsymbol{\alpha})\,\eta\,y_{t-1}(\boldsymbol{\alpha}) + \cdots + \theta_{2q}(\boldsymbol{\alpha})\,\eta^{2q}\,y_{t-2q}(\boldsymbol{\alpha}) = y_t, \tag{5.1}$$

where $0 < \eta \leq 1$ is a parameter that controls the bandwidth of the filter, and the $\theta_j(\boldsymbol{\alpha})$ are coefficients of the filter, depending on $\boldsymbol{\alpha}$ and being symmetric in the sense that

$$\theta_{2q-j}(\boldsymbol{\alpha}) := \theta_j(\boldsymbol{\alpha}) \qquad (j = 0, 1, \ldots, q)$$

with $\theta_0(\boldsymbol{\alpha}) := 1$. This filter was employed by Dragošević and Stanković (1989) to estimate multiple frequencies with the specific choice of the filter coefficients $\theta_j(\boldsymbol{\alpha}) = \alpha_j$. A similar filter was also used by Kay (1984) for estimating complex sinusoids.

Let us assume that the additive noise $\{\epsilon_t\}$ in (1.3) is white with zero mean and finite fourth moment. It will be shown that in this situation a very simple relationship between $\{\theta_k(\boldsymbol{\alpha})\}$ and $\boldsymbol{\alpha}$ can be established according to (A1) and, by appropriately selecting $\boldsymbol{\alpha}$ in a parameter space, theoretical results obtained in previous sections apply.

5.1 Parametrization

Since the noise $\{\epsilon_t\}$ is assumed to be white, it is easy to verify that the autocovariance function of the filtered noise satisfies the equation

$$\sum_{k=0}^{2q} \eta^k\,\theta_k(\boldsymbol{\alpha})\,r^{\epsilon}_{\tau-k}(\boldsymbol{\alpha}) = 0 \qquad (\tau = 1, 2, \ldots), \tag{5.2}$$

or, in matrix form,

$$\bar{\mathbf{r}}_{\epsilon}(\boldsymbol{\alpha}) + \eta^{2q}\,\bar{\mathbf{r}}^{B}_{\epsilon}(\boldsymbol{\alpha}) = -\bar{\mathbf{R}}_{\epsilon}(\boldsymbol{\alpha})\,\tilde{\mathbf{Q}}\,\boldsymbol{\theta}(\boldsymbol{\alpha}). \tag{5.3}$$

In the expression,

$$\boldsymbol{\theta}(\boldsymbol{\alpha}) := \begin{bmatrix} \theta_1(\boldsymbol{\alpha}) \\ \vdots \\ \theta_q(\boldsymbol{\alpha}) \end{bmatrix} \qquad \tilde{\mathbf{Q}} := \begin{bmatrix} \mathbf{Q}_1 & \mathbf{0} \\ \mathbf{0}^T & \eta^q \\ \mathbf{Q}_2 & \mathbf{0} \end{bmatrix}$$

with $\mathbf{Q}_1 := \mathrm{diag}(\eta, \ldots, \eta^{q-1})$ and $\mathbf{Q}_2 := \eta^q \mathbf{Q}_1 \tilde{\mathbf{I}}$. By prior-multiplying each side of (5.3) with $2\,\mathbf{Q}^T$, we obtain

$$(1 + \eta^{2q})\, \mathrm{r}_\epsilon(\boldsymbol{\alpha}) = -2\, \mathbf{Q}^T \bar{\mathbf{R}}_\epsilon(\boldsymbol{\alpha})\, \tilde{\mathbf{Q}}\, \boldsymbol{\theta}(\boldsymbol{\alpha}).$$

Therefore, (A1) requires that

$$\boldsymbol{\theta}(\boldsymbol{\alpha}) = \tfrac{1}{2}(1 + \eta^{2q}) \{\mathbf{Q}^T \bar{\mathbf{R}}_\epsilon(\boldsymbol{\alpha})\, \tilde{\mathbf{Q}}\}^{-1} \mathbf{R}_\epsilon(\boldsymbol{\alpha})\, \boldsymbol{\alpha}. \tag{5.4}$$

On the other hand, simple algebra shows that

$$\mathbf{Q}^T \bar{\mathbf{R}}_\epsilon(\boldsymbol{\alpha})\, \tilde{\mathbf{Q}} = \frac{2}{1 + \eta^{2q}}\, \mathbf{Q}^T \bar{\mathbf{R}}_\epsilon(\boldsymbol{\alpha})\, \mathbf{Q}\, \mathbf{T}_\eta^{-1} = \frac{2}{1 + \eta^{2q}}\, \mathbf{R}_\epsilon(\boldsymbol{\alpha})\, \mathbf{T}_\eta^{-1},$$

where $\mathbf{T}_\eta$ is a q-by-q diagonal matrix of the form

$$\mathbf{T}_\eta := \mathrm{diag}\left(\frac{1 + \eta^{2q}}{\eta + \eta^{2q-1}}, \ldots, \frac{1 + \eta^{2q}}{\eta^{q-1} + \eta^{q+1}}, \frac{1 + \eta^{2q}}{2\,\eta^q} \right).$$

As a result, (5.4) is simplified to a trivial linear equation

$$\boldsymbol{\theta}(\boldsymbol{\alpha}) = \mathbf{T}_\eta\, \boldsymbol{\alpha}. \tag{5.5}$$

With the coefficients given by (5.5), the AR filter in (5.1) satisfies (A1), so that a sequence of estimators $\{\hat{\boldsymbol{\alpha}}_m\}$ can be produced according to FPI in (2.6).

As compared to (5.5), the parametrization in the generalized least squares (GLS) method of Dragošević and Stanković (1989) is given by

$$\boldsymbol{\theta}(\boldsymbol{\alpha}) = \boldsymbol{\alpha}. \tag{5.6}$$

Notice that for $\eta < 1$, (5.5) differs from (5.6). It is this difference that makes the PF estimator consistent for *any* $\eta < 1$, while the GLS estimator is inconsistent. Note also

that PF and GLS coincide when $\eta = 1$. The iterative filtering algorithm (IFA) proposed by Kay (1984) corresponds to the complex version of PFAR with $\eta = 1$. However, Kay used the Burg estimator instead of the least squares in order to guarantee the stability of the filtering.

In the special case of $q = 1$, (5.5) reduces to

$$\boldsymbol{\theta}(\boldsymbol{\alpha}) = \frac{1+\eta^2}{2\eta}\,\boldsymbol{\alpha}.$$

By this choice, (5.1) becomes a variant of the AR(2) filter discussed by Li and Kedem (1991) and Li, Kedem, and Yakowitz (1991). Furthermore, if $\eta = 1$, it coincides with the procedure proposed by Quinn and Fernandes (1991) and Truong-Van (1990).

5.2 Parameter Space

In order for the AR filter to possess the properties (A3) and (A4), the parameter $\boldsymbol{\alpha}$ cannot be arbitrarily. Instead, it must lie inside a parameter space. A possible choice of such a parameter space is presented as follows.

Let Θ be the collection of $\boldsymbol{\theta} := [\theta_1, \cdots, \theta_q]^T$ for which all zeros of the polynomial $\sum_{k=0}^{2q} \theta_k\, z^{2q-k}$ are distinct and on the unit circle in the complex domain, that is,

$$\sum_{k=0}^{2q} \theta_k\, z^{2q-k} = \prod_{k=1}^{q} (z - e^{i\lambda_k})(z - e^{-i\lambda_k})$$

for some $0 < \lambda_1 < \cdots < \lambda_q < \pi$, where $\theta_0 := 1$ and $\theta_{2q-k} := \theta_k$, $(k = 0, \ldots, q-1)$. Define the *parameter space* $\mathcal{A}(\eta)$ to be the collection of $\boldsymbol{\alpha}$ such that $\boldsymbol{\theta}(\boldsymbol{\alpha}) = \mathbf{T}_\eta\, \boldsymbol{\alpha} \in \Theta$, that is,

$$\mathcal{A}(\eta) := \{\boldsymbol{\alpha} :\ \mathbf{T}_\eta\, \boldsymbol{\alpha} \in \Theta\}.$$

Since λ_j take on values in an open region, both Θ and $\mathcal{A}(\eta)$ are also *open* in the q-dimensional space. Moreover, it is readily shown that the poles of the AR filter with $\boldsymbol{\theta}(\boldsymbol{\alpha})$ given by (5.5) occur on the circle $|z| = \eta$ if $\boldsymbol{\alpha} \in \mathcal{A}(\eta)$. Therefore, with $\eta < 1$ and $\boldsymbol{\alpha}^* \in \mathcal{A}(\eta)$, the AR filter satisfies (A3) and (A4) for some bounded and closed subset $\mathcal{A} \subset \mathcal{A}(\eta)$ that contains $\boldsymbol{\alpha}^*$ in its interior.

In practice, it is not guaranteed that $\hat{\boldsymbol{\alpha}}_m$ will always appear inside $\mathcal{A}(\eta)$. In case of falling outside, $\hat{\boldsymbol{\alpha}}_m$ has to be projected back into $\mathcal{A}(\eta)$ in order to satisfy the conditions of the PF method. Suppose that for a given $\boldsymbol{\alpha}$, the zeros of the polynomial $\sum_{k=0}^{2q} \theta_k(\boldsymbol{\alpha})\, z^{2q-k}$ are of the form $\rho_k \exp(\pm i\lambda_k)$, $(k = 1, \ldots, q)$, for some $\rho_k > 0$ and $0 < \lambda_1 < \cdots < \lambda_q < \pi$, where $\{\theta_k(\boldsymbol{\alpha})\}$ are defined by (5.5). Then, the projection of $\boldsymbol{\theta}(\boldsymbol{\alpha})$ on Θ is obviously $\tilde{\boldsymbol{\theta}}(\boldsymbol{\alpha}) := [\tilde{\theta}_1(\boldsymbol{\alpha}), \cdots, \tilde{\theta}_q(\boldsymbol{\alpha})]^T$, where $\tilde{\theta}_k(\boldsymbol{\alpha})$ are determined by the coefficients of the polynomial

$$\sum_{k=0}^{2q} \tilde{\theta}_k(\boldsymbol{\alpha})\, z^{2q-k} = \prod_{k=1}^{q} (z - e^{i\lambda_k})(z - e^{-i\lambda_k}).$$

By this operation, we simply force the poles of the AR filter to occur on the circle $|z| = \eta$. As a result, the projection of $\boldsymbol{\alpha}$ on $\mathcal{A}(\eta)$, denoted by $\tilde{\boldsymbol{\alpha}}$, can be written as

$$\tilde{\boldsymbol{\alpha}} = \mathbf{T}_\eta^{-1}\, \tilde{\boldsymbol{\theta}}(\boldsymbol{\alpha}) \in \mathcal{A}(\eta).$$

5.3 Statistical Properties

For the properties in Section 3 to hold, (A2) is the only thing left for verification. It is readily shown that the AR filter satisfies this condition, since its transfer function

$$H(\omega; \boldsymbol{\alpha}) = \left\{ \sum_{k=0}^{2q} \theta_k(\boldsymbol{\alpha})\, \eta^k e^{-ik\omega} \right\}^{-1}$$

is nonzero for all $\boldsymbol{\alpha} \in \mathcal{A}(\eta)$.

In conclusion, the AR filter, defined by (5.1) and (5.5) with $\eta < 1$, satisfies (A1)–(A4). Consequently, the corresponding PF estimator $\hat{\boldsymbol{\alpha}}$, called the PFAR estimator, possesses the statistical properties in Section 3, regarding the existence, convergence, strong consistency, and asymptotic normality, as can be summarized as follows.

Theorem 5.1 *Suppose that* $\{\epsilon_t\}$ *is white with finite fourth moment. Then, for the AR filter defined by* (5.1) *and* (5.5) *with* $\eta < 1$, *the results in Theorem* 3.1, *Theorem* 3.2, *and Theorem* 3.3 *are valid, provided that* $\boldsymbol{\alpha}^* \in \mathcal{A}(\eta)$.

As aforementioned, the parameter η plays the role of controlling the bandwidth of the AR filter. Indced, the closer η is to 1, the narrower is the bandwidth of the AR filter. Since the sinusoidal signal under investigation concentrates only on extremely narrow bands (spikes) in the frequency domain, it is clear that in order to enhance the signal by the AR filter, η should be chosen as close to 1 as possible. From another point of view, the parameter η also determines the asymptotic behavior of the associated PFAR estimator in terms of its covariance matrix. As a matter of fact, the asymptotic covariance matrix $\mathbf{V}_\alpha$ of the PFAR estimator can be made arbitrarily small if η is chosen arbitrarily close to 1. More precisely, it can be shown that $\mathbf{V}_\alpha$ tends to a zero matrix at the rate of order $(1-\eta)^3$ as $\eta \to 1$.

To verify this assertion, we first need to rewrite $w_{ij}(\boldsymbol{\alpha}^*)$ defined by (3.8) into a more suitable form, with the help of the following spectral representation of the autocovariance function $r_\tau^\epsilon(\boldsymbol{\alpha}^*)$:

$$r_\tau^\epsilon(\boldsymbol{\alpha}^*) = \frac{\sigma_\epsilon^2}{2\pi}\int_{-\pi}^{\pi} |H(\omega;\boldsymbol{\alpha}^*)|^2\, e^{i\tau\omega}\, d\omega.$$

To do so, we notice that (5.2) can be rewritten as

$$\sum_{k=0}^{2q} d_k\, a_k\, r_{\tau-k}^\epsilon(\boldsymbol{\alpha}^*) = 0 \qquad (\tau = 1, 2, \dots),$$

with d_k defined by

$$d_k := \eta^k \theta_k(\boldsymbol{\alpha})/a_k = (\eta^k + \eta^{2q+k})/(\eta^k + \eta^{2q-k}).$$

It turns out, by using the spectral representation of $r_\tau^\epsilon(\boldsymbol{\alpha}^*)$, that

$$\begin{aligned} S_\tau &:= \sum_{k=0}^{2q} a_k\, r_{\tau-k}^\epsilon(\boldsymbol{\alpha}^*) = \sum_{k=0}^{2q}(1-d_k)\, a_k\, r_{\tau-k}^\epsilon(\boldsymbol{\alpha}^*) \\ &= \frac{\sigma_\epsilon^2}{2\pi}\int_{-\pi}^{\pi} |H(\omega;\boldsymbol{\alpha}^*)|^2 \left\{\sum_{k=0}^{2q}(1-d_k)\, a_k\, e^{-ik\omega}\right\} e^{i\tau\omega}\, d\omega. \end{aligned}$$

Introducing the following polynomials

$$D(z) := \sum_{k=0}^{2q}(1-d_k)\, a_k\, z^{2q-k}, \quad P(z) := \sum_{k=0}^{2q} \theta_k(\boldsymbol{\alpha}^*)\, \eta^k\, z^k,$$

and $Q(z) := z^{2q}P(z^{-1})$, we can write S_τ as a Cauchy integral of the form

$$S_\tau = \frac{\sigma_\epsilon^2}{2\pi i} \oint_{|z|=1} \frac{D(z)}{P(z)\,Q(z)}\, z^{\tau-1}\, dz$$

for any $\tau \geq 1$. Note that $\boldsymbol{\alpha}^* \in \mathcal{A}(\eta)$ implies $\boldsymbol{\theta}(\boldsymbol{\alpha}^*) \in \Theta$. This, in turn, guarantees that all of the $2q$ zeros of $Q(z)$ appear on the circle $|z| = \eta$ and can be expressed as $\nu_j := \eta \exp(i\lambda_j)$, with λ_j satisfying $0 < \lambda_1 < \cdots < \lambda_q < \pi$ and $\lambda_{2q-j+1} := -\lambda_j$ for $j = 1, \ldots, q$. As a result, we can write

$$Q(z) = \prod_{j=1}^{2q}(z - \nu_j) \quad \text{and} \quad P(z) = \prod_{j=1}^{2q}(1 - \bar{\nu}_j z),$$

where $\bar{\nu}_j := \eta \exp(-i\lambda_j)$. It follows from the residue theorem of complex analysis that

$$S_\tau = \sigma_\epsilon^2 \sum_{k=1}^{2q} \frac{D(\nu_k)}{P(\nu_k)\,Q'(\nu_k)}\, \nu_k^{\tau-1} \qquad (\tau = 1, 2, \ldots).$$

Using this formula, $w_{ij}(\boldsymbol{\alpha}^*)$ in (3.8) can be written as

$$w_{ij}(\boldsymbol{\alpha}^*) = 4 \sum_{\tau=0}^{\infty} \bar{S}_{\tau+i} S_{\tau+j} = 4 \sum_{k,l=1}^{2q} \frac{\sigma_\epsilon^4\, \overline{D(\nu_k)}\, D(\nu_l)}{\overline{P(\nu_k)}\, P(\nu_l)\, \overline{Q'(\nu_k)}\, Q'(\nu_l)} \frac{\bar{\nu}_k^{i-1} \nu_l^{j-1}}{1 - \bar{\nu}_k \nu_l} \tag{5.7}$$

where the overline denotes the complex conjugation.

Now, with this expression, the behavior of $\mathbf{W}(\boldsymbol{\alpha}^*)$ as η tends to 1 become easer to investigate. In fact, we first notice that $\nu_k \to z_k$ as $\eta \to 1$. Moreover, it is easy to see from the definition of d_j that $d_0 = 1$ and

$$(1 - \eta^2)^{-1}\,(1 - d_j) \to j/2 \qquad (j = 1, \ldots, 2q)$$

as $\eta \to 1$. It follows that

$$(1 - \eta^2)^{-1}\, D(\nu_k) \to \sum_{j=1}^{2q} \tfrac{1}{2} j a_j\, z_k^{2q-j} = -\tfrac{1}{2}\, z_k^{-1} A'(z_k), \tag{5.8}$$

where $A(z)$ is the polynomial given by (1.6). In addition, it is readily shown that

$$(1 - \eta^2)^{-1}\, P(\nu_k) = \prod_{\substack{j=1 \\ j \neq k}}^{2q} (1 - \bar{\nu}_j \nu_k) \to P_k := \prod_{\substack{j=1 \\ j \neq k}}^{2q} (1 - \bar{z}_j z_k)$$

and

$$Q'(\nu_k) = \prod_{\substack{j=1 \\ j\neq k}}^{2q} (\nu_k - \nu_j) \rightarrow \prod_{\substack{j=1 \\ j\neq k}}^{2q} (z_k - z_j) = A'(z_k).$$

Finally, it is easy to see that $(1-\eta^2)/(1-\bar{\nu}_k\nu_l) = 1$ for $l = k$, and $(1-\eta^2)/(1-\bar{\nu}_k\nu_l) \rightarrow 0$ as $\eta \rightarrow 1$ for $l \neq k$. Substituting all these limits in (5.7) yields

$$(1-\eta^2)\, w_{ij}(\boldsymbol{\alpha}^*) \rightarrow \sigma_\epsilon^4 \sum_{k=1}^{2q} \frac{1}{|P_k|^2}\, \bar{z}_k^{i-1} z_k^{j-1}$$

as $\eta \rightarrow 1$. Upon noting that $|P_k| = |A'(z_k)|$, we can also write

$$(1-\eta^2)\, \mathbf{Q}^T \mathbf{W}(\boldsymbol{\alpha}^*)\, \mathbf{Q} \rightarrow 4\sigma_\epsilon^4\, \boldsymbol{\Sigma}_0 \quad \text{with} \quad \boldsymbol{\Sigma}_0 := \mathbf{Q}^T \mathbf{S} \mathbf{D}_0 \mathbf{S}^H \mathbf{Q}, \tag{5.9}$$

where $\mathbf{D}_0$ is a $2q$-by-$2q$ diagonal matrix of the form

$$\mathbf{D}_0 := \frac{1}{4}\, \mathrm{diag}\left\{ \frac{1}{|A'(z_1)|^2}, \ldots, \frac{1}{|A'(z_{2q})|^2} \right\}$$

and $\mathbf{S}$ the Vandermonde matrix in (2.11).

To complete the analysis, we only need to know the behavior of $\mathbf{R}_x(\boldsymbol{\alpha}^*)$ as η tends to 1. For this purpose, we notice that $|H(\omega_k; \boldsymbol{\alpha}^*)|^2 = 1/|Q(z_k)|^2$ and $Q(z_k) = -D(z_k)$. It follows from (5.8) that

$$(1-\eta^2)^2\, |H(\omega_k; \boldsymbol{\alpha}^*)|^2 \rightarrow 4/|A'(z_k)|^2.$$

According to (2.10), this implies that

$$(1-\eta^2)^2\, \mathbf{R}_x(\boldsymbol{\alpha}^*) \rightarrow 8\, r_0^x\, \boldsymbol{\Sigma} \quad \text{with} \quad \boldsymbol{\Sigma} := \mathbf{Q}^T \mathbf{S} \mathbf{D} \mathbf{S}^H \mathbf{Q}. \tag{5.10}$$

In this expression,

$$\mathbf{D} := \frac{1}{4}\, \mathrm{diag}\left\{ \frac{\sigma_1^2}{|A'(z_1)|^2}, \ldots, \frac{\sigma_{2q}^2}{|A'(z_{2q})|^2} \right\},$$

$\sigma_k^2 := \beta_k^2 / \sum_{j=1}^q \beta_j^2$, and $\sigma_{2q-k+1}^2 := \sigma_k^2$, $(k = 1, \ldots, q)$. Introducing the notation $\gamma := r_0^x / r_0^\epsilon = r_0^x / \sigma_\epsilon^2$ as being the signal-to-noise ratio of the data, and substituting all these limits in the expression of $\mathbf{V}_\alpha$ given by (3.12), we finally obtain

$$\lim_{\eta \rightarrow 1} \left(\frac{1+\eta^2}{1-\eta^2} \right)^3 \mathbf{V}_\alpha = \frac{1}{2\gamma^2}\, \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}^{-1}. \tag{5.11}$$

In particular, when the q sinusoidal compoments have the same power, that is, $\beta_k \equiv \beta_1$, (5.11) reduces to

$$\lim_{\eta \to 1} \left(\frac{1+\eta^2}{1-\eta^2} \right)^3 \mathbf{V}_\alpha = \frac{q^2}{2\gamma^2} \boldsymbol{\Sigma}_0^{-1}, \tag{5.12}$$

since in this case $\mathbf{D} = q^{-1}\mathbf{D}_0$.

REMARK. It is interesting to recognize, by comparing (5.9) with (2.10), that $\boldsymbol{\Sigma}_0$ is identical to the autocovariance matrix of a sinusoidal signal having the same q frequencies as $\{x_t\}$, with the amplitude corresponding to ω_k being equal to $1/|A'(z_k)|$. Similarly, $\boldsymbol{\Sigma}$ in (5.10) has the same structure as $\boldsymbol{\Sigma}_0$, but the amplitude associated with ω_k is replaced by $\sigma_k/|A'(z_k)|$. Therefore, these matrices can be rewritten as

$$\boldsymbol{\Sigma}_0 = \left[\sum_{k=1}^{q} \frac{\cos\{\omega_k(i-j)\}}{2\,|A'(z_k)|^2} \right] \quad \text{and} \quad \boldsymbol{\Sigma} = \left[\sum_{k=1}^{q} \frac{\sigma_k^2 \cos\{\omega_k(i-j)\}}{2\,|A'(z_k)|^2} \right]$$

where $i, j = 1, \ldots, 2q-1$.

5.4 Accuracy of the PFAR Estimator

As guaranteed by Theorem 3.3, the PF method produces an estimator $\hat{\boldsymbol{\alpha}}$ so that $\sqrt{n}\,(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)$ is asymptotically normally distributed and thus its estimation accuracy is of order $O(n^{-1/2})$. Upon using the AR filter, it has been shown that the asymptotic covariance matrix of the resulting PF estimator (i.e., the PFAR estimator) can be made arbitrarily small as η tends to 1. This indicates that a higher order accuracy could be obtained with $\eta \approx 1$. Indeed, as discussed in Quinn and Fernandes (1991), Truong-Van (1990), and Li, Kedem, Yakowitz (1991) for the case of a single sinusoid, the PFAR estimator is able to achieve the same accuracy of order $O(n^{-3/2})$ as the nonlinear least squares (NLS) approach (Hannan, 1973; Stoica and Nehorai, 1989; Walker, 1971) in the limiting case of $\eta = 1$.

An advantage of the PF method over the NLS lies in its computational simplicity inherited from the explicit LS solution of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$. Another advantage of the PF method is its less stringent requirement of the initial estimates. As pointed out by many researchers (Rice and Rosenblatt, 1988; Stoica, *et al.*, 1989; Li, Kedem, and Yakowitz, 1991), the

NLS approach, as well as the PFAR estimator in the limiting case of $\eta = 1$, requires an initial estimate of accuracy $o(n^{-1})$, in order to obtain the optimal solution by iterative procedures. On the other hand, an initial estimate of accuracy $O(1)$ is sufficient for the calculation of the PFAR estimator with $\eta < 1$ using FPI in (2.6), as indicated by Theorem 3.1 and also verified by simulations of Li, Kedem, and Yakowitz (1991). More importantly, thanks to the flexibility for the choice of η, the estimation accuracy of the PFAR estimator can be improved significantly upon using a monotone increasing sequence of η such that $0 < \eta_1 < \eta_2 < \cdots \to 1$.

According to the suggestions of Li, Kedem, and Yakowitz (1991), the iteration in (2.6) can be carried out with $\eta = \eta_k$ until convergence, using the PFAR estimator previously obtained with $\eta = \eta_{k-1}$ as the initial value. In so doing for $k = 1, 2, \ldots$, a sequence of PFAR estimators associated with $\{\eta_k\}$ is produced, each estimator in the sequence serving as the initial value of its successor. As k grows, the accuracy improves without requiring a stringent initial estimate. Another possible way of improving the estimation accuracy is to increase η after each iteration rather than carrying on the iteration until convergence (see, e.g., Dragošević and Stanković, 1989; Kedem and Yakowitz, 1991). This strategy simplifies the computation but may result in converging to a false location if η is increased too fast.

In the most interesting case of closely-spaced frequencies, however, one should be very cautious when increasing η. Simulations show (see next section) that the bias of the PF estimator from finite data increases as η approaches 1, while the variance decreases. For closely-spaced frequencies, the bias eventually dominates the variance, and hence an appropriate $\eta < 1$ should be used to balance the trade-off between the bias and variance in minimizing the mean-squared error.

For the case of multiple sinusoids, the fixed-point iteration in (2.6) with $\eta \approx 1$ can be shown to be an algorithm that approximately calculates the NLS estimator in an iterative fashion. In fact, since the NLS estimator minimizes the sum of squared errors

$$J := \sum_{t=1}^{n} \left\{ y_t - \sum_{k=1}^{q} (\hat{A}_k \cos \hat{\omega}_k t + \hat{B}_k \sin \hat{\omega}_k t) \right\}^2 \tag{5.13}$$

with respect to $\hat{A}_k$, $\hat{B}_k$, and $\hat{\omega}_k$, it can be implemented in two steps, similar to the procedure of Bresler and Macovski (1986) for complex sinusoids. The first step is to minimize J with respect to $[\hat{A}_1, \cdots, \hat{A}_q, \hat{B}_1, \cdots, \hat{B}_q]^T$, while keeping $\hat{\omega}_k$ fixed. It is not difficult to verify that the optimal vector is given by $(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\tilde{\mathbf{y}}$, where $\tilde{\mathbf{y}} := [y_1, \cdots, y_n]^T$ and

$$\mathbf{G} := \begin{bmatrix} \cos\hat{\omega}_1 & \cdots & \cos\hat{\omega}_q & \sin\hat{\omega}_1 & \cdots & \sin\hat{\omega}_q \\ \vdots & & \vdots & \vdots & & \vdots \\ \cos n\hat{\omega}_1 & \cdots & \cos n\hat{\omega}_q & \sin n\hat{\omega}_1 & \cdots & \sin n\hat{\omega}_q \end{bmatrix}.$$

Substituting this optimal vector in (5.13) gives

$$J = \|\{\mathbf{I} - \mathbf{G}\,(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\}\tilde{\mathbf{y}}\|^2 = \tilde{\mathbf{y}}^T\{\mathbf{I} - \mathbf{G}\,(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\}\tilde{\mathbf{y}}.$$

This quantity, in turn, is minimized with respect to $\hat{\omega}_k$ in the second step. Notice that $\mathbf{I} - \mathbf{G}\,(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T$ is a projection operator that projects an n-vector onto the orthogonal complement of the $2q$-dimensional column-space of $\mathbf{G}$. On the other hand, if we let $\hat{a}_j$ be the AR parameters determined by (1.6) for any given $\hat{\omega}_k$, and denote by $\hat{\mathbf{A}}$ the corresponding n-by-$(n-2q)$ matrix of the structure (5.18), it is easy to verify that $\hat{\mathbf{A}}^T\mathbf{G} = \mathbf{0}$. This implies that the $n-2q$ linearly-independent columns of $\hat{\mathbf{A}}$ are orthogonal to the columns of $\mathbf{G}$ and thus span the $(n-2q)$-dimensional orthogonal complement of the column-space of $\mathbf{G}$. As a consequence, we obtain

$$\mathbf{I} - \mathbf{G}\,(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T = \hat{\mathbf{A}}\,(\hat{\mathbf{A}}^T\hat{\mathbf{A}})^{-1}\hat{\mathbf{A}}^T$$

and hence $J = \tilde{\mathbf{y}}^T\hat{\mathbf{A}}\,(\hat{\mathbf{A}}^T\hat{\mathbf{A}})^{-1}\hat{\mathbf{A}}^T\tilde{\mathbf{y}}$. Furthermore, let $\hat{\mathbf{e}} := [\hat{e}_{2q+1}, \cdots, \hat{e}_n]^T$, where

$$\hat{e}_t := \sum_{j=0}^{2q} \hat{a}_j\, y_{t-j}. \tag{5.14}$$

Simple algebra shows that $\hat{\mathbf{A}}^T\tilde{\mathbf{y}} = \hat{\mathbf{e}}$. Therefore, J can be rewritten as

$$J = \hat{\mathbf{e}}^T(\hat{\mathbf{A}}^T\hat{\mathbf{A}})^{-1}\hat{\mathbf{e}}. \tag{5.15}$$

The NLS is thus reduced to the problem of minimizing J in (5.15) with respect to $\hat{a}_j$. To compute the NLS estimator, an iterative procedure can be employed in accordance with

the suggestions of Bresler and Macovski (1986). Indeed, for any estimate $\hat{\mathbf{a}}_m$, a matrix $\hat{\mathbf{A}}_m$ can be constructed accordingly and a new estimate can be taken as the minimizer of the criterion $J_m := \hat{\mathbf{e}}^T(\hat{\mathbf{A}}_m^T\hat{\mathbf{A}}_m)^{-1}\hat{\mathbf{e}}$ with respect to $\hat{a}_j$.

Suppose that $\hat{\mathbf{a}}_m := [\hat{a}_1^{(m)}, \cdots, \hat{a}_q^{(m)}]^T$ appears inside a tiny neighborhood of the AR parameter $\mathbf{a}$ that corresponds to the AR model of the signal. It is easy to see that within such a neighborhood $\hat{e}_t$ can be approximated by

$$\hat{e}_t = \sum_{j=0}^{2q} \hat{a}_j^{(m)}\, x_{t-j} + \sum_{j=0}^{2q} \hat{a}_j^{(m)}\, \epsilon_{t-j} \approx \sum_{j=0}^{2q} \hat{a}_j^{(m)}\, \epsilon_{t-j}.$$

Since $\{\epsilon_t\}$ is white, it can be shown from this approximation that the covariance matrix of $\hat{\mathbf{e}}$ is approximately equal to $\sigma_\epsilon^2\hat{\mathbf{A}}_m^T\hat{\mathbf{A}}_m$. Therefore, $\tilde{\mathbf{e}}_m := (\hat{\mathbf{A}}_m^T\hat{\mathbf{A}}_m)^{-1/2}\hat{\mathbf{e}}$ can be regarded as a whitening procedure that decorrelates the vector $\hat{\mathbf{e}}$. Resorting to the frequency-domain interpretation, this whitening procedure can be approximately performed by applying on $\hat{e}_t$ an AR filter of the form (5.1) with $\eta \approx 1$ and $\hat{\mathbf{a}}_m$. Let $\{\tilde{e}_t^{(m)}\}$ be the output of the filter, then $\tilde{\mathbf{e}}_m \approx [\hat{e}_{2q+1}^{(m)}, \cdots, \hat{e}_n^{(m)}]^T$ and hence $J_m = \|\tilde{\mathbf{e}}_m\|^2 \approx \sum(\tilde{e}_t^{(m)})^2$. On the other hand, by interchanging the order of the AR filtering on $\hat{e}_t$ and the FIR filtering on y_t defined by (5.14), we obtain

$$\tilde{e}_t^{(m)} \approx \sum_{j=0}^{2q} \hat{a}_j\, \hat{y}_{t-j}(\hat{\mathbf{a}}_m).$$

Thus, minimizing J_m is approximately equivalent to minimizing $\sum\{\sum_{j=0}^{2q} \hat{a}_j\, \hat{y}_{t-j}(\hat{\mathbf{a}}_m)\}^2$, which yields the PFAR estimator $\hat{\mathbf{a}}_{m+1}$ produced by FPI in (2.6).

The above discussion indicates that for multiple sinusoids the PFAR estimator with η tending 1 is also capable of approaching the accuracy of the NLS procedure, just like the case of a single sinusoid as discussed by Li, Kedem, and Yakowitz (1991).

5.5 The Case of Two Sinusoids

The simplest case of the PFAR estimator for a single sinusoid is similar to the CM estimator discussed by Li and Kedem (1991) and Li, Kedem, and Yakowitz (1991). From their results, it is easy to verify that for $q = 1$ the asymptotic variance $\mathbf{V}_\alpha$ of the PFAR

estimator $\hat{\boldsymbol{\alpha}}$ can be expressed *exactly* by (5.12) without taking the limit, that is,

$$\mathbf{V}_\alpha = \left(\frac{1-\eta^2}{1+\eta^2}\right)^3 \frac{4\sin^2\omega_1}{\gamma^2}.$$

The requirement that $\boldsymbol{\alpha}^* \in \mathcal{A}(\eta)$ reduces to $|\cos\omega_1| < 2\eta/(1+\eta^2)$. This implies that ω_1 should stay away from 0 and π at least by an amount of $\arccos\{2\eta/(1+\eta^2)\}$.

Let us now consider specifically the case of two sinusoids ($q = 2$) in additive white noise. In this case, Θ consists of all $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$ with

$$\theta_1 = -2(\cos\lambda_1 + \cos\lambda_2) \quad \text{and} \quad \theta_2 = 2(1 + 2\cos\lambda_1\cos\lambda_2) \tag{5.16}$$

for some $0 < \lambda_1 < \lambda_2 < \pi$. Simple algebra shows that $\cos\lambda_1$ and $\cos\lambda_2$ must be of the form $\{-\theta_1 \pm (\theta_1^2 - 4\theta_2 + 8)^{1/2}\}/4$. Therefore, θ_1 and θ_2 should satisfy

$$\tfrac{1}{4}\left|-\theta_1 \pm \sqrt{\theta_1^2 - 4\theta_2 + 8}\right| < 1 \quad \text{and} \quad \theta_1^2 - 4\theta_2 + 8 > 0.$$

It turns out by solving these inequalities that Θ is the region defined by

$$\Theta: \quad 2|\theta_1| - 2 < \theta_2 < \tfrac{1}{4}\theta_1^2 + 2. \tag{5.17}$$

Moreover, for $q = 2$, (5.5) reduces to

$$\theta_1(\boldsymbol{\alpha}) = \frac{1+\eta^4}{\eta+\eta^3}\alpha_1 \quad \text{and} \quad \theta_2(\boldsymbol{\alpha}) = \frac{1+\eta^4}{2\eta^2}\alpha_2.$$

According to (5.17), the parameter space $\mathcal{A}(\eta)$ is given by

$$\mathcal{A}(\eta): \quad \frac{4\eta}{1+\eta^2}|\alpha_1| - \frac{4\eta^2}{1+\eta^4} < \alpha_2 < \frac{1+\eta^4}{2(1+\eta^2)^2}\alpha_1^2 + \frac{4\eta^2}{1+\eta^4}. \tag{5.18}$$

Figure 1 shows $\mathcal{A}(\eta)$ for $\eta = 0.8$ together with Θ defined by (5.17).

It is readily seen from Figure 1, as well as (5.18) and (5.5), that $\mathcal{A}(\eta)$ is contained in Θ for $\eta < 1$, and that $\mathcal{A}(\eta)$ coincides with Θ as $\eta \to 1$. Therefore, for any given ω_1 and ω_2 satisfying $0 < \omega_1 < \omega_2 < \pi$, the requirement $\boldsymbol{\alpha}^* \in \mathcal{A}(\eta)$ in Theorem 5.1 can always be met by choosing η close enough to 1.

On the other hand, for a given $\eta < 1$, the requirement $\boldsymbol{\alpha}^* \in \mathcal{A}(\eta)$ imposes a *separation condition* on the frequencies of the signal. As a matter of fact, in order that $\boldsymbol{\alpha}^* \in \mathcal{A}(\eta)$,

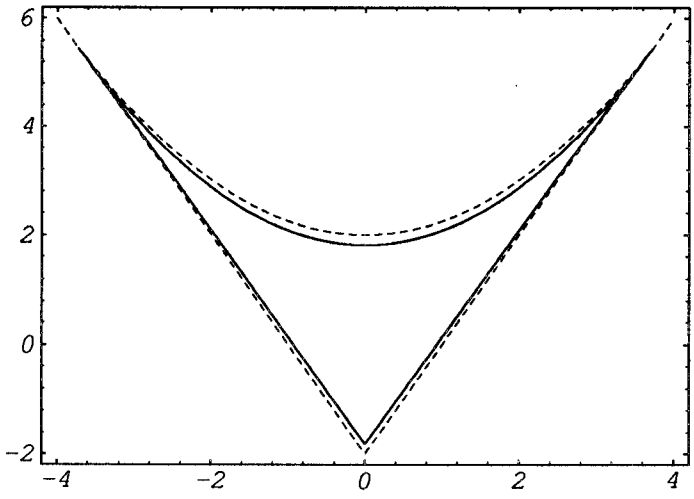


Figure 1: Parameter space $\mathcal{A}(\eta)$ with $\eta = 0.8$ for the case of two sinusoids. The dotted lines define the region Θ.

the frequencies ω_1 and ω_2 should stay away from 0 and π, respectively, and from each other, by a certain amount depending on η. It is not too difficult to verify that a sufficient condition for (5.18) to be fulfilled is that

$$\omega_1,\ \pi - \omega_2 > \arccos\left\{\frac{1+\eta^2}{\sqrt{2(1+\eta^4)}}\right\} \quad \text{and} \quad \omega_2 - \omega_1 > \arccos\left\{\frac{2\eta^2}{1+\eta^4}\right\}.$$

To get a complete picture of the separation condition, Figure 2 shows for $\eta = 0.8$ and 0.9 the set Ω_η, referred to as the *frequency space*, of all frequencies for which $\boldsymbol{\alpha}^* \in \mathcal{A}(\eta)$.

Let us now consider the characteristic of the AR filter at $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$, which determines the asymptotic behavior of the PFAR estimator, as indicated by Theorem 3.3. Let us first look at the squared gain function $|H(\omega; \boldsymbol{\alpha}^*)|^2$. In Figure 3, $|H(\omega; \boldsymbol{\alpha}^*)|^2$ is plotted for $\eta = 0.92$, where $\boldsymbol{\alpha}^* = \mathbf{a}$ is the AR parameter corresponding to the sinusoidal signal with $\omega_1 = 0.32\pi$ and $\omega_2 = 0.45\pi$. It is interesting to observe that the squared gain function has peaks near, but not exactly at, the frequencies of the signal. Therefore, the PFAR method can eventually enhance the signal, while making the estimator consistent by slightly biased peaks. A further illustration of this point is presented in Figure 4, where the poles of the AR filter are plotted against the frequencies of the signal. It can

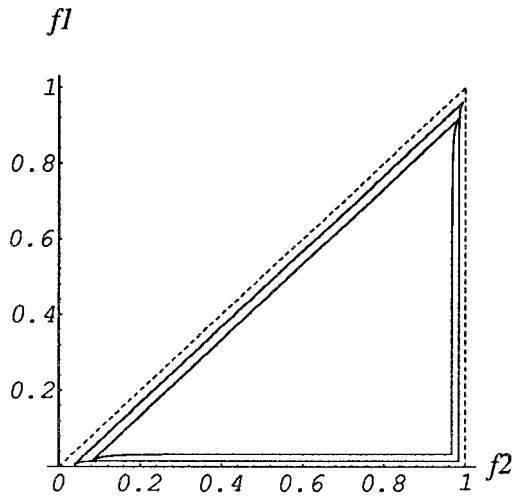


Figure 2: Frequency space of normalized frequency $f := \omega/\pi$ for $\eta = 0.8$ (smallest) and 0.9. The region defined by dotted lines corresponds to the extreme case of $\eta = 1$.

be seen that the PFAR method does not in general force the poles of the AR filter to have the same angular frequencies as those of the signal. Instead, a slightly biased poles are used on the basis of (5.5) in order to produce a consistent estimator, as guaranteed by Theorem 5.1.

6 Experimental Results

In order to illustrate the performance of the PF method, we employ the all-pole filter (5.1) and consider the case of two sinusoids ($q = 2$) in Gaussian white noise. All of the simulations in this section are based on 100 independent realizations of $\{y_t\}$ with a relatively short length of $n = 100$. The phases of the sinusoids are fixed at zero, and the sample variance of the noise is adjusted in each realization according to the sample variance of the signal in order to achieve the required signal-to-noise ratio. Furthermore, in both PF and GLS — which correspond to the parametrizations (5.5) and (5.6) respectively — the poles of the all-pole filter are constrained to be on the circle $|z| = \eta$, by projection if necessary, so that the bandwidth parameter η effectively controls the band-

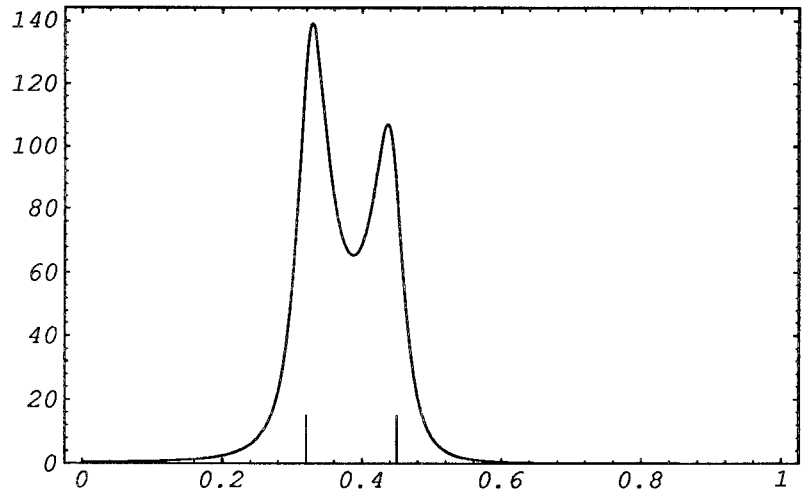


Figure 3: Plot of squared gain with $\eta = 0.92$ and $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ for the case of two sinusoids with $\omega_1 = 0.32\pi$ and $\omega_2 = 0.45\pi$.

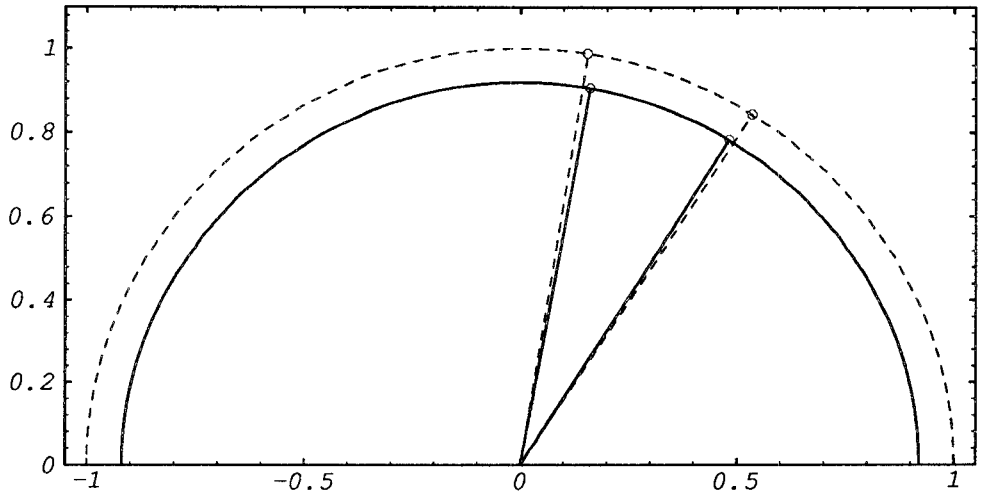


Figure 4: Location of poles of the AR filter shown in Figure 3. The dotted lines indicate frequencies of the signal on the unit circle.

Table 1: PF & GLS Estimates for Well-Separated Frequencies

	PF				GLS			
η	mse	bias	var	complexity	mse	bias	var	complexity
Prony	4.72e-3	4.40e-3	3.23e-4	—	4.72e-3	4.40e-3	3.23e-4	—
0.950	2.20e-6	1.17e-8	2.19e-6	8.0 ± 0.6	2.34e-6	1.48e-7	2.19e-6	8.1 ± 0.8
0.960	1.96e-6	9.14e-9	1.96e-6	3.1 ± 0.2	2.04e-6	9.47e-8	1.95e-6	3.2 ± 0.4
0.970	1.79e-6	9.72e-9	1.78e-6	3.2 ± 0.3	1.84e-6	6.21e-8	1.77e-6	3.3 ± 0.4
0.980	1.67e-6	1.64e-8	1.65e-6	3.3 ± 0.4	1.69e-6	4.96e-8	1.65e-6	3.4 ± 0.4
0.985	1.63e-6	2.39e-8	1.60e-6	2.9 ± 0.4	1.65e-6	4.79e-8	1.60e-6	2.8 ± 0.4
0.990	1.59e-6	3.60e-8	1.55e-6	3.0 ± 0.5	1.60e-6	5.08e-8	1.55e-6	2.7 ± 0.4
0.995	1.55e-6	5.53e-8	1.49e-6	3.4 ± 0.4	1.55e-6	6.08e-8	1.49e-6	2.9 ± 0.6
1.000	1.50e-6	8.27e-8	1.42e-6	3.7 ± 0.6	1.51e-6	8.29e-8	1.42e-6	3.4 ± 0.5

width of the all-pole filter and the performance of the estimators. For convenience, the following simulation results are given in regard to the *normalized* frequencies $f_k := \omega_k/\pi$, and the average mean-squared error

$$\text{mse} := \tfrac{1}{2}\{E(\hat{f}_1 - f_1)^2 + E(\hat{f}_2 - f_2)^2\}$$

is employed as an overall performance index. Moreover, we define the average bias and average variance of the frequency estimates by

$$\text{bias} := \tfrac{1}{2}\{(E(\hat{f}_1) - f_1)^2 + (E(\hat{f}_2) - f_2)^2\} \quad \text{and} \quad \text{var} := \tfrac{1}{2}\{\text{var}(\hat{f}_1) + \text{var}(\hat{f}_2)\}$$

respectively. The frequency estimates of both PF and GLS are obtained by the fixed-point iteration (2.6) in connection with (5.16) which provides the relationship between the frequency and AR estimates. The stopping rule of the iteration is given by

$$\sqrt{(\hat{f}_1^{(m)} - \hat{f}_1^{(m-1)})^2 + (\hat{f}_2^{(m)} - \hat{f}_2^{(m-1)})^2} < 10^{-5}.$$

In other words, the iteration terminates at the mth iteration if this inequality is satisfied.

Table 2: PF & GLS Estimates for Closely-Spaced Frequencies

	PF				GLS			
η	mse	bias	var	complexity	mse	bias	var	complexity
Prony	1.46e-2	1.40e-2	5.68e-4	—	1.46e-2	1.40e-2	5.68e-4	—
0.950	1.81e-6	2.69e-8	1.78e-6	10.6 ± 3.4	4.68e-6	3.05e-6	1.63e-6	11.0 ± 5.3
0.960	1.67e-6	6.20e-8	1.61e-6	3.3 ± 0.4	3.63e-6	2.17e-6	1.46e-6	3.8 ± 0.2
0.970	1.63e-6	1.63e-7	1.46e-6	3.8 ± 0.3	2.95e-6	1.59e-6	1.36e-6	3.9 ± 0.2
0.980	1.75e-6	4.21e-7	1.32e-6	4.3 ± 0.3	2.56e-6	1.29e-6	1.27e-6	3.7 ± 0.4
0.985	1.90e-6	6.57e-7	1.25e-6	4.2 ± 0.4	2.46e-6	1.25e-6	1.21e-6	2.8 ± 0.4
0.990	2.15e-6	9.83e-7	1.16e-6	4.4 ± 0.3	2.45e-6	1.31e-6	1.15e-6	3.0 ± 0.5
0.995	2.47e-6	1.41e-6	1.06e-6	4.5 ± 0.5	2.56e-6	1.50e-6	1.06e-6	3.7 ± 0.7
1.000	2.85e-6	1.91e-6	9.45e-7	4.5 ± 0.8	2.86e-6	1.91e-6	9.46e-7	4.3 ± 0.8

In our simulations, we first compare the performance of PF and GLS in the cases of well-separated and closely-spaced frequencies. In both cases the SNR is fixed at 0 dB per sinusoid, while the bandwidth parameter η in the all-pole filter (5.1) takes different values. Since η varies, it is convenient to explicitly write the corresponding frequency estimates $(\hat{f}_1(\eta), \hat{f}_2(\eta))$ as functions of η. Table 1 and Table 2 present some statistics of the frequency estimates for eight *ascending* values of η, that is, $\eta_1 = 0.95$, $\eta_2 = 0.96, \ldots, \eta_8 = 1$. The mean and variance of the stopping time m are also given as "complexity" in the form of "mean $\pm$ variance".

Notice that in the fixed-point iteration (2.6) two things are needed: a value of the bandwidth parameter η and an initial guess of the AR parameter $\mathbf{a}$. In both PF and GLS, we use Prony's estimator $\mathbf{a}_{\mathrm{LS}}$ as the initial guess of $\mathbf{a}$ corresponding the first value of η, i.e., $\eta = \eta_1 = 0.95$. When the iteration terminates, the resulting AR estimate, denoted by $\hat{\mathbf{a}}(\eta_1)$, is used not only to obtain the frequency estimates $(\hat{f}_1(\eta_1), \hat{f}_2(\eta_1))$, but also to initiate the iteration for the next value of η, namely, $\eta = \eta_2 = 0.96$. In general, as η grows from η_2 to η_8, we employ the previous AR estimate $\hat{\mathbf{a}}(\eta_{k-1})$ to initiate the

iteration (2.6) corresponding to $\eta = \eta_k$ in the computation of $\hat{\mathbf{a}}(\eta_k)$.

In Table 1, the true frequencies were taken to be $(f_1, f_2) = (0.41, 0.59)$, which are well-separated as compared to the width of a Fourier bin $\Delta f = 2/n = 0.02$ in our simulations. As we can see, Prony's estimator gives poor frequency estimates, while both PF and GLS significantly improve Prony's estimator in terms of mean-squared error, even with a relatively small η. Moreover, the estimation accuracy can be further improved by increasing η toward 1. This is quite intuitive because increasing η is equivalent to shrinking the bandwidth of the all-pole filter and thereby enhancing the sinusoids (see also Dragošević and Stanković, 1989; Kedem and Yakowitz, 1991) for similar ideas of shrinking bandwidth in frequency estimation). The issue is that the shrinkage should be based upon reliable frequency estimates, since otherwise the filter may lock on false locations (an example in this regard will be presented later). This is the reason why in our simulations the iteration for a higher value of η was initiated by the estimates corresponding to a slightly smaller value of η. In this way, we gradually shrink the bandwidth as more reliable estimates become available. Table 1 shows that as η approaches 1 the PF and GLS estimates achieve a precision (mse) of 1.50×10^{-6} — very close to the asymptotic variance of the nonlinear least squares (NLS) estimator which in this case equals 1.22×10^{-6}. Thus, for well-separated frequencies, PF and GLS have the same final performance as η approaches 1.

For closely-spaced frequencies, however, the PF estimator performs better than GLS, as can be seen from Table 2 and Figure 5. In this experiment, the true frequencies were taken to be $(f_1, f_2) = (0.47, 0.51)$ while all other conditions remained the same as in the previous example. Notice that the true frequencies are separated only by two Fourier bins. It is interesting to observe that as the bandwidth parameter η increases toward 1 the mse of both methods does not monotonically decrease as in the case of well-separated frequencies. Instead, it starts increasing after a certain value of η (see also Figure 5). The reason is the following. A closer examination of Table 1 and Table 2 reveals that as η approaches 1 the bias increases while the variance decreases in both methods. For well-separated frequencies as shown in Table 1, the bias never dominates the variance,

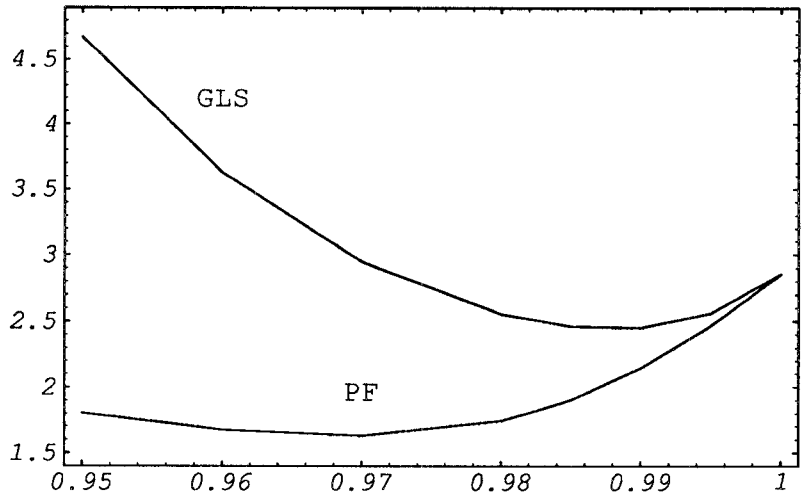


Figure 5: Plot of mse ($\times 10^{-6}$) against η for closely-spaced frequencies.

Table 3: Estimation With $\eta = 1$

(f_1, f_2)	mse	$E(\hat{f}_1) \pm \text{var}(\hat{f}_1)$	$E(\hat{f}_2) \pm \text{var}(\hat{f}_2)$	complexity
(0.41,0.59)	4.33e-6	$0.409704 \pm 1.30\text{e-}6$	$0.590529 \pm 6.99\text{e-}6$	17.9 ± 10.7
(0.47,0.51)	1.38e-4	$0.468628 \pm 9.37\text{e-}7$	$0.513672 \pm 2.49\text{e-}4$	24.3 ± 63.3

and hence the mse decreases basically along with the decrease of the variance. On the other hand, the bias becomes dominant in the case of closely-spaced frequencies as η approaches 1 (see Table 2), and a trade-off effect between bias and variance takes place. As we can see from Figure 5, the best value of η for the PF estimator lies between 0.96 and 0.98 where the mse achieves the smallest values. The GLS estimator is clearly inferior to the PF estimator in this example because of its relatively higher bias. Indeed, the bias and variance of the GLS estimator play an equal role in the mse, since their magnitudes are of the same order (see Table 2).

Table 2 illustrates the role of η as a parameter that can be utilized to balance the bias and variance of the PF estimator for minimizing the mean-squared error. Now, in Table 3, we illustrate the role of η in the convergence of the fixed-point iteration (2.6)

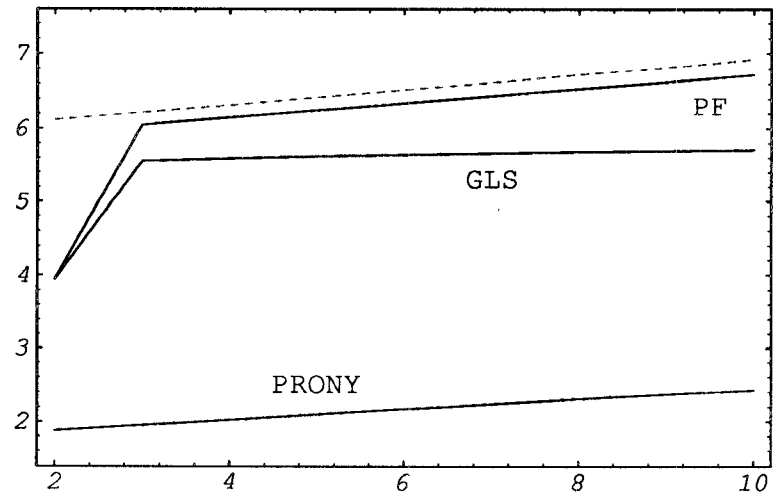


Figure 6: Plot of $-\log(\text{mse})$ against SNR in dB for very closely-spaced frequencies. The dotted line indicates the asymptotic variance of NLS.

when initial guesses are poor. Instead of gradually increasing η toward 1, as done in Table 1 and Table 2, the frequency estimates in Table 3 were obtained right away with $\eta = 1$, using Prony's estimator as the initial guess. This is equivalent to the iterative procedure that employs the all-pole filter *without* η and starts with Prony's estimator. As can be seen from Table 3, the mean-squared error is higher than the mse reported in Table 1 and Table 2 corresponding to (gradually achieved) $\eta = 1$, especially for closely-spaced frequencies. This indicates that without η in the all-pole filter the iteration (2.6) may fail to converge to the desired fixed-point when poor initial guesses, such as Prony's estimator, are used. The reason is that the bandwidth of the all-pole filter without η (or, equivalently, with $\eta = 1$) is extremely narrow. Although it could be helpful to have a narrow bandwidth for enhancement of the sinusoids if good initial guesses are used, a narrow bandwidth might cause the filter to "loose" the true frequencies when tuned according to inaccurate estimates. Therefore, the safest way is to start with a relatively small η, to accommodate even poor initial guesses, and then gradually increase η as improved estimates from previous iterations become available.

To show the performance of the PF estimator under different signal-to-noise ratios,

Figure 6 presents the negative logarithm of the mse for various values of SNR, with the dotted line indicating the asymptotic variance of NLS as a reference. In this example, the frequencies were taken to be $(f_1, f_2) = (0.485, 0.495)$, which are closely-spaced *within* a Fourier bin, and the bandwidth parameter η was fixed at 0.985 in both PF and GLS. Prony's estimator again was used to initiate the fixed-point iteration (2.6) for both methods. As can be seen, the mse of the PF estimator closely follows the asymptotic variance of NLS when SNR $\geq$ 3 dB, and the performance of both PF and GLS deteriorates rapidly when the SNR is below this threshold. The poor initial accuracy of Prony's estimator is largely responsible for this particular value of threshold. In fact, simulations show that the threshold can be extended to -2 dB if the initial guesses are taken to be the two Fourier frequencies which correspond to the largest absolute values in the FFT of the data.

In Figure 7, we show that the PF estimator is capable of resolving very closely-spaced frequencies where the GLS method fails. In this experiment, the true frequencies, $(f_1, f_2) = (0.41, 0.412)$, are only 10% apart relative to the width of a Fourier bin. The frequency estimates were obtained with $\eta = 0.997$ and the fixed-point iteration (2.6) was initiated by Prony's estimator. Figure 7(a) shows the negative logarithm of the mse for different values of SNR and Figure 7(b) presents the averages of the frequency estimates. Compared to the GLS estimator, the PF estimator has a much smaller bias which enables it to resolve the frequencies as well as to achieve a smaller mean-squared error. Notice that the GLS estimator gives essentially a single frequency $f \approx 0.411$ between the two true frequencies.

Finally, we note that in the preceding discussion the phases were fixed at zero. Experience shows, however, that when the phases are chosen at random the mse may worsen somewhat. This is understandable due to the small sample size which cannot explain the addition of extra sources of variability (Kay and Marple, 1981).

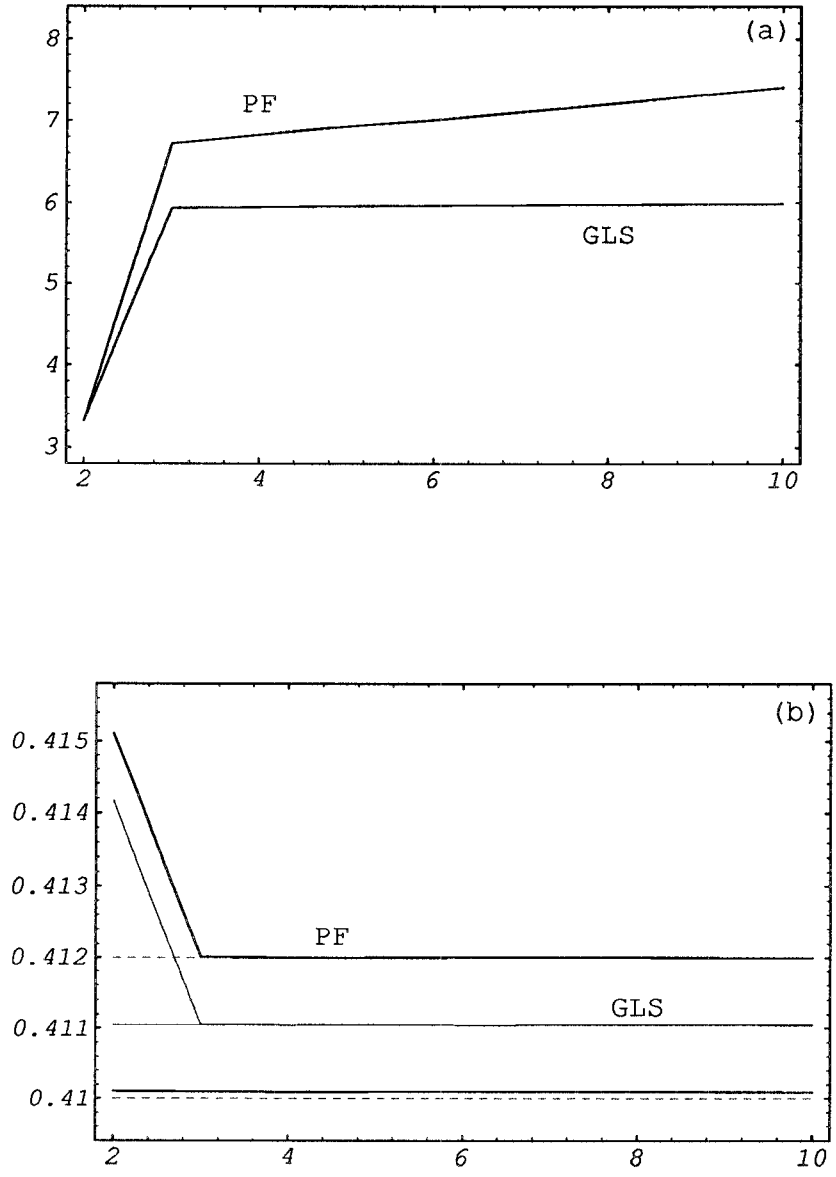


Figure 7: (a) Plot of $-\log(\text{mse})$ against SNR in dB. (b) Plot of averaged frequency estimates against SNR in dB with dotted lines indicating true frequencies.

References

[1] Bresler, Y. and A. Macovski (1986). Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1081–1089.

[2] Brockwell, P. J. and R. A. Davis (1987). *Time Series: Theory and Methods.* New York: Springer-Verlag.

[3] Dragošević, M. V. and S. S. Stanković (1989). A generalized least squares method for frequency estimation. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 6, pp. 805–819.

[4] Hannan, E. J. (1973). The estimation of frequency. *J. Appl. Prob.*, vol. 10, pp. 510–519.

[5] He, S. and B. Kedem (1990). Higher-order crossings of an almost periodic random sequence in noise. *IEEE Trans. Inform. Theory*, vol. 35, pp. 360–370.

[6] Kay, S. M. (1984). Accurate frequency estimation at low signal-to-noise ratio. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 3, pp. 540–547.

[7] Kay, S. M. and S. L. Marple (1981). Spectrum analysis – a modern perspective. *Proc. of IEEE*, vol. 69, no. 11, pp. 1380–1419.

[8] Lehmann, E. L. (1983). *Theory of Point Estimation.* New York: Wiley.

[9] Kedem, B. and S. Yakowitz (1991). Practical aspects of a fast algorithm for frequency detection. Submitted for publication.

[10] Li, T. H. and B. Kedem (1991). Strong consistency of the contraction mapping method for frequency estimation. Submitted.

[11] Li, T. H., B. Kedem, and S. Yakowitz (1991). Asymptotic normality of the contraction mapping estimator for frequency estimation. Submitted.

[12] Ortega, J. M. and W. C. Rheinboldt (1970). *Iterative Solution of Nonlinear Equations in Several Variables.* New York: Academic.

[13] Quinn, B. G. and J. M. Fernandes (1991). A fast efficient technique for the estimation of frequency. *Biometrika*, vol. 78, no. 3, pp. 489–497.

[14] Rice, J. A. and M. Rosenblatt (1988). On frequency estimation. *Biometrika*, vol. 75, no. 3, pp. 477–484.

[15] Stoer, J. and R. Bulirsch (1980). *Introduction to Numerical Analysis.* New York: Springer-Verlag.

[16] Stoica, P., R. L. Moses, B. Friedlander, and T. Söderström (1989). Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 3, pp. 378–392.

[17] Stoica, P. and A. Nehorai (1989). Statistical analysis of two nonlinear least-squares estimators of sine-wave parameters in the colored-noise case. *Circuits, Systems, Signal Processing*, vol. 8, no. 1, pp. 3–15.

[18] Truong-Van, B. (1990). A new approach to frequency analysis with amplified harmonics. *J. R. Statist. Soc.*, series B, vol. 52, pp. 347–366.

[19] Walker, A. M. (1971). On the estimation of a harmonic component in a time series with stationary independent residuals. *Biometrika*, vol. 58, no. 1, pp. 21–36.

[20] Yakowitz, S. (1991). Some contribution to a frequency location method due to He and Kedem. *IEEE Trans. Inform. Theory*, vol. 37, pp. 1177–1181, 1991.